

154679

STIC-Biotech/ChemLib

From: Whiteman, Brian  
Sent: Friday, May 27, 2005 9:45 AM  
To: STIC-Biotech/ChemLib  
Subject: seq search

09/610,313  
Barnett et al.

SEQ ID NOs: 30, 31, and 32  
1) nucleotide search against databases  
2) amino acid search against databases

Thank you,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

RECEIVED  
MAY 27 2005  
(STIC)

5

\*\*\*\*\*

STAFF USE ONLY

Searcher: \_\_\_\_\_  
Searcher Phone: 2-\_\_\_\_\_  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: 6/3  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search

NA#: \_\_\_\_\_ AA#: \_\_\_\_\_  
Interference: \_\_\_\_\_ SPDI: \_\_\_\_\_  
S/L: \_\_\_\_\_ Oligomer: \_\_\_\_\_  
Encode/Transl: \_\_\_\_\_  
Structure#: \_\_\_\_\_ Text: \_\_\_\_\_  
Inventor: \_\_\_\_\_ Litigation: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1270.42 Seconds  
(without alignments)  
11504.726 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469

Sequence: 1 gtcagccaccatggccga.....gggctagcaccgtgaattc 2469

Scoring table: IDENTITY NUC

Gapop 10'0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2469	100.0	2469	6	ABL39959 Synthetic
2	2469	100.0	2469	12	Adm73764 HIV-1 pol
3	2457	99.5	2457	8	ACA03548 Synthetic
4	2457	99.5	2457	10	ADCI3266 DNA of HI
5	2442.2	98.9	2463	6	ABL39960 Synthetic
6	2442.2	98.9	2463	12	ADMW3765
7	2434.6	98.6	2457	8	ACA03547 Synthetic
8	2434.6	98.6	2457	10	ADCI3265
9	2415.4	97.8	2457	6	ABL39961 Synthetic
10	2415.4	97.8	2457	12	ADMW3766
11	2401.8	97.3	2445	8	ACA03546 Synthetic
12	2401.8	97.3	2445	10	ADCI3264 DNA of HI
13	2394.8	97.0	3930	10	ADCI3230 DNA of HI
14	2393.2	96.9	3930	10	ADCI3231 DNA of HI
15	2393.2	96.9	3930	10	ADCI3232 DNA of HI
16	2393.2	96.9	5184	8	ACA03591 Synthetic
17	2393.2	96.9	5184	10	ADCI3279
18	2362.8	95.7	3531	10	ADCI3234 DNA of HI
19	2361.2	95.6	3537	10	ADCI3236 DNA of HI
20	2360.2	95.6	5145	8	ACA03521 Synthetic

21	2360.2	95.6	5145	10	ADCI3233	Adci3233 DNA of HI
22	2350.2	95.2	3538	10	ADCI3235	Adci3235 DNA of HI
23	2349.4	95.2	3624	8	ACA03550	ACA03550 Synthetic
24	2349.4	95.2	3624	10	ADCI3268	Adci3268 DNA of HI
25	2301.8	93.2	3607	8	ACA03551	ACA03551 Synthetic
26	2301.8	93.2	3607	10	ADCI3269	Adci3269 DNA of HI
27	2283.6	92.5	3597	8	ACA03549	ACA03549 Synthetic
28	2283.6	92.5	3597	10	ADCI3267	Adci3267 DNA of HI
29	2149.6	87.1	2472	8	ACA03543	ACA03543 Synthetic
30	2149.6	87.1	2472	8	ACC78507	Acc78507 HIV p2Pol
31	2121.2	85.9	2466	8	ACA03542	ACA03542 Synthetic
32	2121.2	85.9	2466	8	ACC78506	Acc78506 HIV p2Pol
33	2094.4	84.8	2460	8	ACA03541	ACA03541 Synthetic
34	2094.4	84.8	2460	8	ACC78505	Acc78505 HIV p2Pol
35	2093.4	84.8	3564	8	ACC78488	Acc78488 HIV GagPo
36	2093.4	84.8	3564	8	ACC78489	Acc78489 HIV GagPo
37	2092.8	84.8	4716	8	ACA03522	ACA03522 Synthetic
38	2092.8	84.8	4716	10	ADCI3238	Adci3238 DNA of HI
39	2089	84.6	3999	8	ACC78484	Acc78484 HIV GagCo
40	2087.4	84.5	3999	8	ACC78485	Acc78485 HIV GagCo
41	2087.4	84.5	3999	8	ACC78486	Acc78486 HIV GagCo
42	2087.4	84.5	5283	8	ACA03584	ACA03584 Synthetic
43	2087.4	84.5	5283	8	ACC78529	Acc78529 HIV TatRe
44	2087.2	84.5	4713	8	ACA03592	ACA03592 Synthetic
45	2087.2	84.5	4713	10	ADCI3280	Adci3280 DNA of HI

## ALIGNMENTS

### RESULT 1

ABL39959  
ID ABL39959 standard; DNA; 2469 BP.  
XX  
AC ABL39959;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Synthetic construct PR975(+) SEQ ID NO:30.  
XX  
KW Human immunodeficiency virus type C; antigenic HIV type C protein; immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef; immunostimulant; Gene therapy; gene; ds.  
XX  
OS Human immunodeficiency virus; type C.  
XX  
PN WO200204493-A2.  
XX  
PD 17-JAN-2002.  
XX  
PF 05-JUL-2001; 2001WO-US021241.  
XX  
PR 05-JUL-2000; 2000US-00610313.  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;  
XX  
DR WPI; 2002-154920/20.  
XX  
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.  
XX  
PS Claim 1; Fig 8; 233pp; English.  
CC The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (i). (i) have immunostimulant activity and can be used in gene therapy.

CC	The HIV type C polynucleotides are useful in applications including DNA									
CC	immunisation, generation of packaging cell lines, and production of HIV									
CC	Type C proteins. The polynucleotides are particularly useful in gene									
CC	therapy and DNA immunisation applications. ABL39942 to ABL40054 and									
CC	ABB06204 to ABB06215 represent sequences used in the exemplification of									
CC	the present invention									
XX										
SQ	Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;									
	Query Match	100.0%; Score 2469; DB 6; Length 2469;								
	Best Local Similarity	100.0%; Pred. No. 7.4e-298;								
	Matches 2469; Conservative	0; Mismatches 0; Indels 0; Gaps 0;								
QY	1	GTGAGCGCCACCATGCGCGAGGCGCATGAGCAGCGCCACAGCGCCCAACATCTCTGATGTCAG	60							
DB	1	GTGAGCGCCACCATGCGCGAGGCGCATGAGCAGCGCCACAGCGCCCAACATCTCTGATGTCAG	60							
QY	61	CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGC	120							
DB	61	CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGC	120							
QY	121	CACATCGCGCGCAACTGCGCGCGCCCGCGCAAGAGGGCTGCTGGAAGTGCAGCAAGGAG	180							
DB	121	CACATCGCGCGCAACTGCGCGCGCCCGCGCAAGAGGGCTGCTGGAAGTGCAGCAAGGAG	180							
QY	181	GGCCACCATGAGGAGTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCC	240							
DB	181	GGCCACCATGAGGAGTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCC	240							
QY	241	TTCCCGCCAGGCGAGGCGCGAGTTTCCCGAGGAGCAGAACCGCGCAACAGCCCCACC	300							
DB	241	TTCCCGCCAGGCGAGGCGCGAGTTTCCCGAGGAGCAGAACCGCGCAACAGCCCCACC	300							
QY	301	AGCGCGAGTGCAGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCAG	360							
DB	301	AGCGCGAGTGCAGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCAG	360							
QY	361	GGCACCTGAACTTCCCGCAGATCACCTCTGTGGCAGCGCCCTTGTGTAGCATCAAGGTG	420							
DB	361	GGCACCTGAACTTCCCGCAGATCACCTCTGTGGCAGCGCCCTTGTGTAGCATCAAGGTG	420							
QY	421	GGCGGCGCAGATCAAGGAGGCGCTGTGTGACACCGCGCGCGACGACACCGTGTGGAGGAG	480							
DB	421	GGCGGCGCAGATCAAGGAGGCGCTGTGTGACACCGCGCGCGACGACACCGTGTGGAGGAG	480							
QY	481	ATGAGCTTGC CGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAG	540							
DB	481	ATGAGCTTGC CGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAG	540							
QY	541	GTGCGCGAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTG	600							
DB	541	GTGCGCGAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTG	600							
QY	601	CTGATCGGCGCCACCGCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGC	660							
DB	601	CTGATCGGCGCCACCGCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGC	660							
QY	661	ACCCTGAACTTCCCGCATCAGCCCCCATCGAGACCGTGC CGGTGAAGCTGAAGCCCGCATG	720							
DB	661	ACCCTGAACTTCCCGCATCAGCCCCCATCGAGACCGTGC CGGTGAAGCTGAAGCCCGCATG	720							
QY	721	GACGGCCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCGCTTACCGCC	780							
DB	721	GACGGCCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCGCTTACCGCC	780							
QY	781	ATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCCCGCGAGAACCCCTAC	840							
DB	781	ATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCGCCCGCGAGAACCCCTAC	840							
QY	841	AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCAGCAGTGGCGCGAGCTGGTGGAC	900							
DB	841	AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCAGCAGTGGCGCGAGCTGGTGGAC	900							

QY	901	TTCCGCGAGTGAACAAGCGCAGCCAGGACTTCTTGGAGAGTGCAGCTGGGCGATCCGCCAC	960						
DB	901	TTCCGCGAGTGAACAAGCGCAGCCAGGACTTCTTGGAGAGTGCAGCTGGGCGATCCGCCAC	960						
QY	961	CCCGCGGCTGAAAGAAAGAGAGCGTGACCGTGTGAGACGTGGGCGAGCGCTACTTC	1020						
DB	961	CCCGCGGCTGAAAGAAAGAGAGCGTGACCGTGTGAGACGTGGGCGAGCGCTACTTC	1020						
QY	1021	AGGTCGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC	1080						
DB	1021	AGGTCGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC	1080						
QY	1081	AACGAGACCCCGGCGCATCCGCTACCAAGTACAAAGTGTGCTGCCAGGGCTGGAAGGGCAGC	1140						
DB	1081	AACGAGACCCCGGCGCATCCGCTACCAAGTACAAAGTGTGCTGCCAGGGCTGGAAGGGCAGC	1140						
QY	1141	CCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCC	1200						
DB	1141	CCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCC	1200						
QY	1201	GAGTCGTGATCTTACCAAGTACATGAGCAGCTGTAGTGGGAGCGACCTTGGAGATCGGC	1260						
DB	1201	GAGTCGTGATCTTACCAAGTACATGAGCAGCTGTAGTGGGAGCGACCTTGGAGATCGGC	1260						
QY	1261	CAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGGCTTCAACACC	1320						
DB	1261	CAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGGCTTCAACACC	1320						
QY	1321	CCGCAAGAGACACAGAAAGAGGAGCCCTTCTGTGTGATGGGCTACGAGCTGCACCCC	1380						
DB	1321	CCGCAAGAGACACAGAAAGAGGAGCCCTTCTGTGTGATGGGCTACGAGCTGCACCCC	1380						
QY	1381	GACAAAGTGACCGTGCAGGCCCTCGAGCTGCGGAGAGGAGAGCTTGGAGCGGAAAGCAG	1440						
DB	1381	GACAAAGTGACCGTGCAGGCCCTCGAGCTGCGGAGAGGAGAGCTTGGAGCGGAAAGCAG	1440						
QY	1441	ATCAGAAAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTTACCCCGGCGATCAAGGTG	1500						
DB	1441	ATCAGAAAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTTACCCCGGCGATCAAGGTG	1500						
QY	1501	CGCAGCTGTGAAGCTGTGCGGCGCGCAAGGCGCTTGAACCGACATCTGTCGCGCTGACCC	1560						
DB	1501	CGCAGCTGTGAAGCTGTGCGGCGCGCAAGGCGCTTGAACCGACATCTGTCGCGCTGACCC	1560						
QY	1561	GAGGAGCGGAGCTGAGCTGCGCGGAGAACCGGAGATCTGCGCGAGCCGCTGCACCGGC	1620						
DB	1561	GAGGAGCGGAGCTGAGCTGCGCGGAGAACCGGAGATCTGCGCGAGCCGCTGCACCGGC	1620						
QY	1621	GTGTACTACGACCCCGCAAGGACCTGTGTGGCCGAGATCCAGAAAGCAGGGCCACGACAG	1680						
DB	1621	GTGTACTACGACCCCGCAAGGACCTGTGTGGCCGAGATCCAGAAAGCAGGGCCACGACAG	1680						
QY	1681	TGACACTACAGATCTTACAGGAGCCCTTCAAGAACTTGAAGACCGGAGTACGCCAAG	1740						
DB	1681	TGACACTACAGATCTTACAGGAGCCCTTCAAGAACTTGAAGACCGGAGTACGCCAAG	1740						
QY	1741	ATCGCACCGCCCAACACACGACGTGAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCC	1800						
DB	1741	ATCGCACCGCCCAACACACGACGTGAAGCAGCTGACCGAGGCGCTGCAAGAGATCGCC	1800						
QY	1801	ATGGAGAGCATCTGTGATCTTGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAG	1860						
DB	1801	ATGGAGAGCATCTGTGATCTTGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAG	1860						
QY	1861	ACCTGGGAGACTGGTGGACCGGACTACTTGGCAGCGCCACTTGGATCCCCGAGTGGAGTTC	1920						
DB	1861	ACCTGGGAGACTGGTGGACCGGACTACTTGGCAGCGCCACTTGGATCCCCGAGTGGAGTTC	1920						
QY	1921	GTCAACACCCCGCTTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1980						
DB	1921	GTCAACACCCCGCTTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1980						
QY	1981	GCGGAGACCTTCTTACGTTGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGCGC	2040						

Db 1981 GCCGAGACCTTCTAGTGGACGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGCGCGGC 2040  
 Qy 2041 TACGTGACCGACCGGGCGGCGAGAGATCGTAGCCTGACCGAGACCAACCAAG 2100  
 Db 2041 TACGTGACCGACCGGGCGGCGAGAGATCGTAGCCTGACCGAGACCAACCAAG 2100  
 Qy 2101 ACCGAGCTGACGAGGCATCAGCTGGCCCTGAGGACACGCGGAGGTTGAACATCGTG 2160  
 Db 2101 ACCGAGCTGACGAGGCATCAGCTGGCCCTGAGGACACGCGGAGGTTGAACATCGTG 2160  
 Qy 2161 ACCGACGACGAGTACGCTGGCCCTGAGGACACGCGGAGGTTGAACATCGTG 2220  
 Db 2161 ACCGACGACGAGTACGCTGGCCCTGAGGACACGCGGAGGTTGAACATCGTG 2220  
 Qy 2221 CTGGTGAACCAAGATCATCAGCAGCTGATCAAGAGGAGAGAGGTGATCTGAGCTGGGTG 2280  
 Db 2221 CTGGTGAACCAAGATCATCAGCAGCTGATCAAGAGGAGAGAGGTGATCTGAGCTGGGTG 2280  
 Qy 2281 CCCGCCCAAGGGGATCGGGCGCAAGAGGATCGAAGCTGGTGAGCAAGGGCATC 2340  
 Db 2281 CCCGCCCAAGGGGATCGGGCGCAAGAGGATCGAAGCTGGTGAGCAAGGGCATC 2340  
 Qy 2341 CGCAAGGTGCTTCTTGGAGCGGATCGATGGCGGCATCGTGATCTACCAATGAGAC 2400  
 Db 2341 CGCAAGGTGCTTCTTGGAGCGGATCGATGGCGGCATCGTGATCTACCAATGAGAC 2400  
 Qy 2401 GACCTGTACGTGGGCGGCGGCGCTAGGATCGATTAAGGCTTCCCGGGCTAGCAC 2460  
 Db 2401 GACCTGTACGTGGGCGGCGGCGCTAGGATCGATTAAGGCTTCCCGGGCTAGCAC 2460  
 Qy 2461 GGTGAATTC 2469  
 Db 2461 GGTGAATTC 2469

RESULT 2

ADMT3764  
 ID ADMT3764 standard; DNA; 2469 BP.  
 AC ADMT3764;  
 XX ADMT3764;  
 DT 03-JUN-2004 (first entry)  
 XX HIV-1 polynucleotide #7.  
 DE HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
 KW HIV type C protein; immunostimulant.  
 XX Human immunodeficiency virus 1.  
 OS US2003223961-A1.  
 PN 04-DEC-2003.  
 PD 05-JUL-2001; 2001US-00899575.  
 XX 05-JUL-2000; 2000US-00610313.  
 XX (MEGE/) MEGEDE J Z.  
 PA (BARN/) BARNETT S W.  
 PA (ENGE/) ENGELBRECHT S.  
 PA (RENS/) RENSBERG E J V.  
 XX Merged JZ, Barnett SW, Engelbrecht S, Rensburg EJV;  
 FI WPI; 2004-060515/06.  
 XX New expression cassette comprising a polynucleotide sequence encoding an  
 XX HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
 PT immunization, generating of packaging cell lines or in producing HIV Type  
 XX C proteins.

PS Claim 1; SEQ ID NO 30; 160pp; English.  
 XX The invention relates to an expression cassette comprising a  
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
 CC also relates to a recombinant expression system for use in a host cell  
 CC comprising an expression cassette, where the polynucleotide sequence  
 CC further comprises control elements capable of driving expression in the  
 CC selected host cell, a cell comprising an expression cassette where the  
 CC polynucleotide sequence further comprises control elements compatible  
 CC with the expression in the cell and a composition for generating an  
 CC immunological response, comprising an expression cassette. The expression  
 CC cassette and the methods of the invention are useful in eliciting an  
 CC immune response, in DNA immunisation, in generation of packaging cell  
 CC lines and in producing HIV Type C proteins. This sequence represents an  
 CC HIV-1 polynucleotide of the invention.  
 XX  
 SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 2469; DB 12; Length 2469;  
 Best Local Similarity 100.0%; Pred. No. 7.4e-298;  
 Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GTCCAGCCACCATGGCGCGGCGCATGAGCCAGGCCACCGGCCAACATCTGTATGCAG 60  
 Db 1 GTCCAGCCACCATGGCGCGGCGCATGAGCCAGGCCACCGGCCAACATCTGTATGCAG 60  
 Qy 61 CGCAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
 Db 61 CGCAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120  
 Qy 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Db 121 CACATCGCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Qy 181 GGCACCCAGATGAAGGACTGCACCGAGCGCCAGCGCCAACTTCTTCCGCGAGACCTGGCC 240  
 Db 181 GGCACCCAGATGAAGGACTGCACCGAGCGCCAGCGCCAACTTCTTCCGCGAGACCTGGCC 240  
 Qy 241 TTCCCCCGAGGCAAGGGCCCGGAGTTCCTCCAGCGAGCAGAAACCGCGCCCAACAGCCCCACC 300  
 Db 241 TTCCCCCGAGGCAAGGGCCCGGAGTTCCTCCAGCGAGCAGAAACCGCGCCCAACAGCCCCACC 300  
 Qy 301 AGCCGCGAGTGCAGGTGGCGGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCAG 360  
 Db 301 AGCCGCGAGTGCAGGTGGCGGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCAG 360  
 Qy 361 GGCACCTGAACTTCCCGCGAGATCACCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
 Db 361 GGCACCTGAACTTCCCGCGAGATCACCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
 Qy 421 GCGCGCCAGATCAAGGAGGCGCTCTGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 Db 421 GCGCGCCAGATCAAGGAGGCGCTCTGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
 Qy 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
 Db 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
 Qy 541 GTGCGCCAGTACGACGAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600  
 Db 541 GTGCGCCAGTACGACGAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600  
 Qy 601 CTGATCGCGCCCG 660  
 Db 601 CTGATCGCGCCCG 660  
 Qy 661 ACCCTGAATTCCTCCCATCAGCCCCCATCGAGACCGCGCGCGCGCGCGCGCGCGCGCG 720  
 Db 661 ACCCTGAATTCCTCCCATCAGCCCCCATCGAGACCGCGCGCGCGCGCGCGCGCGCGCG 720  
 Qy 721 GACGCGCCCAAGGTGAAGCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
 Db 721 GACGCGCCCAAGGTGAAGCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780

```
QY 781 ATCTGCAGGAGTGGAGAGGCGCAAGATCACCAAGATCGSCCCCGAGAACCCCTAC 840
Db 781 ATCTGCAGGAGTGGAGAGGCGCAAGATCACCAAGATCGSCCCCGAGAACCCCTAC 840
QY 841 AACACCCCGCTGTTCGCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGAC 900
Db 841 AACACCCCGCTGTTCGCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGAC 900
QY 901 TTCGGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCGCCAC 960
Db 901 TTCGGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCGCCAC 960
QY 961 CCGCGCGCTGGAAGAAGAAGAGCGTGACCGTCTGGACGTGGCGCAGCGCTACTTC 1020
Db 961 CCGCGCGCTGGAAGAAGAAGAGCGTGACCGTCTGGACGTGGCGCAGCGCTACTTC 1020
QY 1021 AGCGTGCCCTGGACGAGGACTTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080
Db 1021 AGCGTGCCCTGGACGAGGACTTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080
QY 1081 AACGAGACCCCGGCATCCGCTACCAAGTACACGCTGCTGCCCGCAGGCGTGGAAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCATCCGCTACCAAGTACACGCTGCTGCCCGCAGGCGTGGAAAGGGCAGC 1140
QY 1141 CCAGAGCATCTTCAGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCGCCGCAACCCC 1200
Db 1141 CCAGAGCATCTTCAGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCGCCGCAACCCC 1200
QY 1201 GAGATCGTGATCTACCAAGTACATGGACGACCTGTGCTGGCGCAGCGACCTGGAGATCGGC 1260
Db 1201 GAGATCGTGATCTACCAAGTACATGGACGACCTGTGCTGGCGCAGCGACCTGGAGATCGGC 1260
QY 1261 CAGCACCGCGCAAGTTCGAGAGCTGCGCAAGCACTGTGCTGGCGTGGGGCTTCAACACC 1320
Db 1261 CAGCACCGCGCAAGTTCGAGAGCTGCGCAAGCACTGTGCTGGCGTGGGGCTTCAACACC 1320
QY 1321 CCGACAAGAAGCACCAAGAGGAGCCCTTCTGTGGATGGCTACGAGCTGCACCCC 1380
Db 1321 CCGACAAGAAGCACCAAGAGGAGCCCTTCTGTGGATGGCTACGAGCTGCACCCC 1380
QY 1381 GACAAGTGGACCGGTGACGCCATCGAGCTGCCCGAGAGAGCTGGAACCGTGAACGAC 1440
Db 1381 GACAAGTGGACCGGTGACGCCATCGAGCTGCCCGAGAGAGCTGGAACCGTGAACGAC 1440
QY 1441 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG 1500
Db 1441 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG 1500
QY 1501 CGCAGCTGTGAAGCTGTGCGCGGCGCAAGGCCCTGACGACATCGTGCCCTGACC 1560
Db 1501 CGCAGCTGTGAAGCTGTGCGCGGCGCAAGGCCCTGACGACATCGTGCCCTGACC 1560
QY 1561 GAGGAGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGCAACGGC 1620
Db 1561 GAGGAGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGCAACGGC 1620
QY 1621 GTGTACTACGCCCGCAGCAAGGACTGTGGTGGCGAGATCCAGAAGCAAGGCGCCACGACAG 1680
Db 1621 GTGTACTACGCCCGCAGCAAGGACTGTGGTGGCGAGATCCAGAAGCAAGGCGCCACGACAG 1680
QY 1681 TGGACCTTACCAGATCTTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1740
Db 1681 TGGACCTTACCAGATCTTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1740
QY 1741 ATGGCACCCGCCACACCAAGCAGTGAAGCAGTGAACCGGCGCTGCGAGAGATCGCC 1800
Db 1741 ATGGCACCCGCCACACCAAGCAGTGAAGCAGTGAACCGGCGCTGCGAGAGATCGCC 1800
QY 1801 ATGGAGCATGTGATCTGGGGCAGACCCCAAGTTCGCGCTGCCATCCAGAGAGGAG 1860
Db 1801 ATGGAGCATGTGATCTGGGGCAGACCCCAAGTTCGCGCTGCCATCCAGAGAGGAG 1860
```

```
QY 1861 ACCTGGGAGACCTGTGGAGCCGACTACTTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Db 1861 ACCTGGGAGACCTGTGGAGCCGACTACTTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
QY 1921 GTGAAACACCCCGCTGTGTGAAGCTGTGTGTAACAGTGGAGAGAGCCCATCATCGGC 1980
Db 1921 GTGAAACACCCCGCTGTGTGAAGCTGTGTGTAACAGTGGAGAGAGCCCATCATCGGC 1980
QY 1981 GCCGAGACCTTCTACCTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2040
Db 1981 GCCGAGACCTTCTACCTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2040
QY 2041 TAGCTGACCCGACCGGGCGCCGCGCAGAAAGATCGTGAGCTGACCGAGACCAACCCAGAAG 2100
Db 2041 TAGCTGACCCGACCGGGCGCCGCGCAGAAAGATCGTGAGCTGACCGAGACCAACCCAGAAG 2100
QY 2101 ACCGAGTGCAGGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGGTTGAACATCGTG 2160
Db 2101 ACCGAGTGCAGGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGGTTGAACATCGTG 2160
QY 2161 ACCGAGCAGCAGTACCGCCCTGGGCGCATCTCCAGGCCAGCCGACAGAGAGAGCGAG 2220
Db 2161 ACCGAGCAGCAGTACCGCCCTGGGCGCATCTCCAGGCCAGCCGACAGAGAGAGCGAG 2220
QY 2221 CTGCTGAACCAAGGATTCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTG 2280
Db 2221 CTGCTGAACCAAGGATTCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTG 2280
QY 2281 CCGCCCGACAAAGGCGATCGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATC 2340
Db 2281 CCGCCCGACAAAGGCGATCGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATC 2340
QY 2341 CGCAAGTGTCTTCTGCGACGCGATCGATGGCGCATCGTGATCTACCAAGTACATGGAC 2400
Db 2341 CGCAAGTGTCTTCTGCGACGCGATCGATGGCGCATCGTGATCTACCAAGTACATGGAC 2400
QY 2401 GACTGTACGTGGCGAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2460
Db 2401 GACTGTACGTGGCGAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2460
QY 2461 GGTGAATTC 2469
Db 2461 GGTGAATTC 2469
RESULT 3
ACA03548
ID ACA03548 standard; DNA; 2457 BP.
XX
AC ACA03548;
XX
AC ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-031680P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
```



```
Db 1681 TACCAGATCTACCAAGAGCCCTTTCAGAACTGTAAGACCGCAAGTACGCAAGATCGCG 1740
Qy 1747 ACCGCCACACCAACGACGTGAAGACGCTACCGAGGCCCTGTCAGAGATCGCCATGGAG 1806
Db 1741 ACCGCCACACCAACGACGTGAAGACGCTACCGAGGCCCTGTCAGAGATCGCCATGGAG 1800
Qy 1807 AGCATCGTGTCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGGAGACCTGG 1866
Db 1801 AGCATCGTGTCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGGACCTGG 1860
Qy 1867 GAGACCTGGTGGACCGCACTACTGCGAGGCCACCTGGATCCCGAGTGGGATTCGTGAAC 1926
Db 1861 GAGACCTGGTGGACCGCACTACTGCGAGGCCACCTGGATCCCGAGTGGGATTCGTGAAC 1920
Qy 1927 ACCCCCCCTTGTGAAGCTGTGGTACCACTGACAGTGGAGAGGCCCATCATCGCGGCCGAG 1986
Db 1921 ACCCCCCCTTGTGAAGCTGTGGTACCACTGAGAGAGGAGGCCCATCATCGCGGCCGAG 1980
Qy 1987 ACCTTCTACGTGGACGGCGCGCCCAACCGGAGACCAAGATCGCAAGCGCGGTACGTG 2046
Db 1981 ACCTTCTACGTGGACGGCGCGCCCAACCGGAGACCAAGATCGCAAGCGCGGTACGTG 2040
Qy 2047 ACCGACCGGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2106
Db 2041 ACCGACCGGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2107 CTGAGGCCATCAGCTGGGCCCTTCAGGACAGCGGCGAGGTGAACATCGTGACCGAC 2166
Db 2101 CTGAGGCCATCAGCTGGGCCCTTCAGGACAGCGGCGAGGTGAACATCGTGACCGAC 2160
Qy 2167 AGCAGTACGCCCTTGGGCATCATCCAGGCCAGCCGACCAAGAGCGAGCGAGCTGGTG 2226
Db 2161 AGCAGTACGCCCTTGGGCATCATCCAGGCCAGCCGACCAAGAGCGAGCGAGCTGGTG 2220
Qy 2227 AACGAGATCATGAGCAGCTGTATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCC 2286
Db 2221 AACGAGATCATGAGCAGCTGTATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy 2287 CACAGGGCATCGCGGCAAGCAGCAGATCGACAGCTGTGTGAGCAAGGGCATCCGCAAG 2346
Db 2281 CACAGGGCATCGCGGCAAGCAGCAGATCGACAGCTGTGTGAGCAAGGGCATCCGCAAG 2340
Qy 2347 GTGCTGTTCCTGGAACGCGATCGATGGCGGATCGTGATCTACAGTACATGACGACCTG 2406
Db 2341 GTGCTGTTCCTGGAACGCGATCGATGGCGGATCGTGATCTACAGTACATGAGACCTG 2400
Qy 2407 TAGTGGGACAGCGCGCCCTTAGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db 2401 TAGTGGGACAGCGCGCGCCCTTAGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457

RESULT 4
ADCI3266
ID ADCI3266 standard; DNA; 2457 BP.
AC AC
AC ADCI3266;
XX AC
XX ADCI3266 (first entry)
XX DT.
XX 18-DEC-2003
XX DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX DE
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX Human immunodeficiency virus.
XX OS
XX WO2003004620-A2.
XX PN
XX 16-JAN-2003.
XX PD
XX 05-JUL-2002; 2002WO-US021420.
XX PF
XX 05-JUL-2001; 2001US-0303192P.
XX PR
```

```
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX (CHIR ) CHIRON CORP.
XX (UYST-) UNIV STELLENBOSCH.
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX WPI; 2003-221593/21.
XX New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX Disclosure; Fig 42; 301pp; English.
XX The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramusosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC type C related sequence of the invention.
XX SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;
Query Match 99.5%; Score 2457; DB 10; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2.3e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 7 GCCACCATGCGCGAGGCCATGAGCCAGGCCACCGAGGCCAACATCTCTGATGCGAGCGCAGC 66
Db 1 GCCACCATGCGCGAGGCCATGAGCCAGGCCACCGAGGCCAACATCTCTGATGCGAGCGCAGC 60
Qy 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGGGCCACATC 126
Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGGGCCACATC 120
Qy 127 GCCCGCAACTGCGCGCCCCCGCCCAAGAGGGCTGCTGGAAGTGCAGAGAGGGGCCAC 186
Db 121 GCCCGCAACTGCGCGCCCCCGCCCAAGAGGGCTGCTGGAAGTGCAGAGAGGGGCCAC 180
Qy 187 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACCTTCTTCGCGAGGACCTTGGCTTCCCC 246
Db 181 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACCTTCTTCGCGAGGACCTTGGCTTCCCC 240
Qy 247 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAACCGGCCCAACAGCCCCCAGCGCCG 306
Db 241 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAACCGGCCCAACAGCCCCCAGCGCCG 300
Qy 307 GAGCTCAGGTGCGCGGCAACACACCCCGAGGAGCGCGGCCCGCGAGCGCCAGGGCACC 366
Db 301 GAGCTCAGGTGCGCGGCAACACACCCCGAGGAGCGCGGCCCGCGAGCGCCAGGGCACC 360
Qy 367 CTGAACCTTCCCCCAGATCACCTTGTGCGAGCGCCCCCTTGTGTAGCATCAAGGTGGCGGC 426
Db 361 CTGAACCTTCCCCCAGATCACCTTGTGCGAGCGCCCCCTTGTGTAGCATCAAGGTGGCGGC 420
Qy 427 CAGATCAAGGAGGCCCTTGTGGACACCGGCGCGCAGACACCTGTCTGGAGGAGATGAGC 486
Db 421 CAGATCAAGGAGGCCCTTGTGGACACCGGCGCGCAGACACCTGTCTGGAGGAGATGAGC 480
Qy 487 CTGCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGCGGCTTTCATCAAGGTGCGC 546
Db 481 CTGCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGCGGCTTTCATCAAGGTGCGC 540
Qy 547 CAGTACGACCGAGATCTCTGATTCGAGATCTGCGGCAAGAGGCCATCGGACACCGTCTGTATC 606
Db 547 CAGTACGACCGAGATCTCTGATTCGAGATCTGCGGCAAGAGGCCATCGGACACCGTCTGTATC 606
```

Db 541 CAGTACGACAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 600  
Qy 607 GGGCCCAACCCCGTGAACATCATCGGCCGCAACATGTGACCCCAAGCTGGGCTGACCCCTG 666  
Db 601 GGGCCCAACCCCGTGAACATCATCGGCCGCAACATGTGACCCCAAGCTGGGCTGACCCCTG 660  
Qy 667 AACTTCCCATCAGCCCATCGAGACCGTGGCCGCTGAAGCTGAAGCCCGGATGACGGC 726  
Db 661 AACTTCCCATCAGCCCATCGAGACCGTGGCCGCTGAAGCTGAAGCCCGGATGACGGC 720  
Qy 727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGC 786  
Db 721 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGC 780  
Qy 787 GAGGAGTGGAGAGAGGGGCAAGATCAACAAGATCGGCCCGGAGAACCCCTTACAACACC 846  
Db 781 GAGGAGTGGAGAGAGGGGCAAGATCAACAAGATCGGCCCGGAGAACCCCTTACAACACC 840  
Qy 847 CCCGTGTTCCGCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGACTTCGGC 906  
Db 841 CCCGTGTTCCGCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGACTTCGGC 900  
Qy 907 GAGCTGAACAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGCACTCCGCCACCCCGCC 966  
Db 901 GAGCTGAACAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGCACTCCGCCACCCCGCC 960  
Qy 967 GGCCTGAAGAGAGAGAGCGTGAACCGTGTGGACGTGGCGGACGCGCTACTTTCAGGGTG 1026  
Db 961 GGCCTGAAGAGAGAGAGCGTGAACCGTGTGGACGTGGCGGACGCGCTACTTTCAGGGTG 1020  
Qy 1027 CCCCTGACGAGGACTTCGCGAAGTACACCGCTTACCATCCCGAGGATCAACAACGAG 1086  
Db 1021 CCCCTGACGAGGACTTCGCGAAGTACACCGCTTACCATCCCGAGGATCAACAACGAG 1080  
Qy 1087 ACCCCCGGCACTCCGCTACAGTACAGCTGCTGCCCGAGGGCTGGAAGGGCAGCCCGAGC 1146  
Db 1081 ACCCCCGGCACTCCGCTACAGTACAGCTGCTGCCCGAGGGCTGGAAGGGCAGCCCGAGC 1140  
Qy 1147 ATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACAGTACATGAGCAGCTGTGTCGGGAGGAGCTGAGATGCGGCAGGAC 1266  
Db 1201 GTGATCTACAGTACATGAGCAGCTGTGTCGGGAGGAGCTGAGATGCGGCAGGAC 1260  
Qy 1267 CGGSCCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTACCAACCCCGGAC 1326  
Db 1261 CGGSCCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTACCAACCCCGGAC 1320  
Qy 1327 AAGAAGCACAGAGGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAG 1386  
Db 1321 AAGAAGCACAGAGGAGCCCGCTTCTGTGGATGGGCTACGAGCTGCACCCCGCAAG 1380  
Qy 1387 TGGACCTGTGACGCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAG 1446  
Db 1381 TGGACCTGTGACGCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAG 1440  
Qy 1447 AAGCTGTGGGCAAGCTGAAGCTGGGCGAGCCAGCATCTACCCCGGCATCAAGGTGCGCAG 1506  
Db 1441 AAGCTGTGGGCAAGCTGAAGCTGGGCGAGCCAGCATCTACCCCGGCATCAAGGTGCGCAG 1500  
Qy 1507 CTGTGAAGCTGTGTCGGGGCGCAAGCCCTGACCGACATCGTGGCCCTGACCGAGGAG 1566  
Db 1501 CTGTGAAGCTGTGTCGGGGCGCAAGCCCTGACCGACATCGTGGCCCTGACCGAGGAG 1560  
Qy 1567 GCCGAGCTGAGCTGGCCGAGAACCCGCGAGATCTGCGCGAGCCCGTGCACCGCGTGTAC 1626  
Db 1561 GCCGAGCTGAGCTGGCCGAGAACCCGCGAGATCTGCGCGAGCCCGTGCACCGCGTGTAC 1620  
Qy 1627 TACGACCCGACAGGAGCTGGTGGCCGAGATCAGAGCAGGGCCAGCAGCAGTGAACC 1686  
Db 1621 TACGACCCGACAGGAGCTGGTGGCCGAGATCAGAGCAGGGCCAGCAGCAGTGAACC 1680

Qy 1687 TACCAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATGGC 1746  
Db 1681 TACCAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATGGC 1740  
Qy 1747 ACCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGCCGTGCAGAACTGCCATGGAG 1806  
Db 1741 ACCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGCCGTGCAGAACTGCCATGGAG 1800  
Qy 1807 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCCATCCAGAGGAGACCTGG 1866  
Db 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCCATCCAGAGGAGACCTGG 1860  
Qy 1867 GAGACTGTGTGAACCGACTACTGTGAGCCCACTTGGATCCCGAGTGGGAGTTTGTGAAC 1926  
Db 1861 GAGACTGTGTGAACCGACTACTGTGAGCCCACTTGGATCCCGAGTGGGAGTTTGTGAAC 1920  
Qy 1927 ACCCCCGCTTGGTGAAGCTGTGTGACAGCTGAGAGGAGCCCATCATCGGCGCCGAG 1986  
Db 1921 ACCCCCGCTTGGTGAAGCTGTGTGACAGCTGAGAGGAGCCCATCATCGGCGCCGAG 1980  
Qy 1987 ACCTTTCTACGTGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACCTGTG 2046  
Db 1981 ACCTTTCTACGTGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACCTGTG 2040  
Qy 2047 ACCGACCGGGCCCGGCAAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAG 2106  
Db 2041 ACCGACCGGGCCCGGCAAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
Qy 2107 CTGCAAGCCATTCAGCTGGGCTTCCAGACAGCGCGAGCGAGGTGAACATCGTGAACGAC 2166  
Db 2101 CTGCAAGCCATTCAGCTGGGCTTCCAGACAGCGCGAGCGAGGTGAACATCGTGAACGAC 2160  
Qy 2167 AGCCAGTACGCGCTTGGGCACTCATCCAGCCCGAGCCGACAGAGCGAGCGAGCTGGTG 2226  
Db 2161 AGCCAGTACGCGCTTGGGCACTCATCCAGCCCGAGCCGACAGAGCGAGCGAGCTGGTG 2220  
Qy 2227 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTAACCTGAGCTGGGTGCCCGCC 2286  
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTAACCTGAGCTGGGTGCCCGCC 2280  
Qy 2287 CACAGGGCATCGGCGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGCATCGGCAAG 2346  
Db 2281 CACAGGGCATCGGCGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGCATCGGCAAG 2340  
Qy 2347 GTGCTGTTCTTGGACGCGCATCGATGGCGGATCTGATCTACAGTACATGACGACCTG 2406  
Db 2341 GTGCTGTTCTTGGACGCGCATCGATGGCGGATCTGATCTACAGTACATGACGACCTG 2400  
Qy 2407 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463  
Db 2401 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

## RESULT 5

ABL39960

ID ABL39960 standard; DNA; 2463 BP.

XX ABL39960;

AC ABL39960;

XX 15-MAY-2002 (first entry)

XX Synthetic construct PR975YM SEQ ID NO:31.

XX Human immunodeficiency virus type C; antigenic HIV type C protein; nef;

XX immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;

XX immunostimulant; gene therapy; gene; ds.

XX Human immunodeficiency virus; type C.

OS Synthetic.

XX WO200204493-A2.

PN

XX



PD	17-JAN-2002.	QY	541	GTGCGCCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGCCATCGCACCGTG	600
XX		DB	541		
XX	05-JUL-2001; 2001WO-US021241.				
XX					
PR	05-JUL-2000; 2000US-00610313.	QY	601	CTGATCGGCCCCACCCCGTGAAACATCATCGGCGGCAACATGCTGACCCAGCTGGGCTGC	660
XX	(CHIR ) CHIRON CORP.	DB	601		
PA	(UYST-) UNIV STELLENBOSCH.				
XX		QY	661	ACCTCGAATCTCCCATCATCAGCCCATCGAGACCGTCCCGTGAGCTGAAGCCCGGATG	720
PI	Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	DB	661		
XX	WPI; 2002-154920/20.				
DR		QY	721	GACGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCC	780
XX	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in	DB	721		
PT	applications including DNA immunization or generation of packaging cell				
PT	lines, particularly in gene therapy.	QY	781	ATCTGCGAGAGATGAGAGAAGGGGCAAGATCACCAGATCGGCCCGGAGAACCCCTAC	840
XX	Claim 1; Fig 9; 233pp; English.	DB	781		
XX		QY	841	AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC	900
CC	The present invention describes expression cassettes comprising a	DB	841		
CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV				
CC	type C polypeptides. The expression cassettes comprise any of the HIV	QY	901	TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCAATCCCCAC	960
CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef	DB	901		
CC	(i) have immunostimulant activity and can be used in gene therapy.				
CC	The HIV type C polynucleotides are useful in applications including DNA	QY	961	CCGCGCGGCTGAGAGAAGAGAGAGCGTGCACCGTCTGGACGTGGCGGCGGCTACTTC	1020
CC	immunisation, generation of packaging cell lines, and production of HIV	DB	961		
CC	Type C proteins. The polynucleotides are particularly useful in gene				
CC	therapy and DNA immunisation applications. ABL39942 to ABL40054 and	QY	1021	AGCGTCCCTTGGACGAGGACTTCCGCAAGTACACCGCCTTCAACCATCCCCAGCATCAAC	1080
CC	ABB06204 to ABB06215 represent sequences used in the exemplification of	DB	1021		
CC	the present invention				
XX	Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;	QY	1081	AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCTGCGGCTGGAAGGGGAGC	1140
SQ		DB	1081		
		QY	1141	CCGAGATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCC	1200
		DB	1141		
		QY	1201	GAGATCGTATCTACCAAGTACATGGACGACCTGTAGTGGGACAGCGACCTGGAGATCGGC	1260
		DB	1201		
		QY	1261	CAGCACCGCGCCAAGATCGAGGAGCTGGCGAAGCACTGTGCTGCTGGGGCTTCAACACC	1320
		DB	1261		
		QY	1321	CCGCAAGAGACCAAGAGGAGGAGCCCTTCTGTGGATGGCTACGAGCTGCACCCC	1380
		DB	1321		
		QY	1381	GACAAAGTGACCGTGCAGCGCCATCGAGCTGCCCGAGAGGAGAGCTGGAACCGTGAACGAC	1440
		DB	1381		
		QY	1441	ATCCAGAAGCTGTGGGCAAGCTGAJCTGGGCGCAGCAGATCTACCCCGGCATCAAGGTG	1500
		DB	1441		
		QY	1501	CGCAGCTGTGAAGCTGTGCGGGCGCAAGGCCCTGACCCACATCGTGCCCTGACCC	1560
		DB	1501		
		QY	1561	GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCCGCGAGATCTGTGGCGAGCCCGTGCAAGGC	1620
		DB	1561		
		QY	1621	GTGTACTACGACCCCAAGAGGACCTGGTGGCGGAGATCCAGAAAGCAAGGCGGCGACGACG	1680
		DB	1621		

Query Match 98.9%; Score 2442.2; DB 6; Length 2463;  
Best Local Similarity 99.6%; Pred. No. 1.6e-294;  
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

1 GTGCGCCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGCCATCGCACCGTG 60  
1 GTGCGCCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGCCATCGCACCGTG 60

61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGC 120  
61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGC 120

121 CACATGCGCGCAACTGCGCGGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180  
121 CACATGCGCGCAACTGCGCGGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180

181 GGCACACAGATGAAGGACTGCACCGAGCGCGAGCCAACTTCTTCCGCGAGGACCTGGCC 240  
181 GGCACACAGATGAAGGACTGCACCGAGCGCGAGCCAACTTCTTCCGCGAGGACCTGGCC 240

241 TTCCCGCAGGCGAGGCCCCGCGAGTTTCCCGAGGAGCAGAACCGCGCAACAGCCCCACC 300  
241 TTCCCGCAGGCGAGGCCCCGCGAGTTTCCCGAGGAGCAGAACCGCGCAACAGCCCCACC 300

301 AGCGCGAGCTGAGGTGCGCGGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGCGAG 360  
301 AGCGCGAGCTGAGGTGCGCGGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGCGAG 360

361 GGCACCTGAATTTCCCGCAGATCACCTGTGGCAGCGGCCCTTGGTGAAGTGAAGGTG 420  
361 GGCACCTGAATTTCCCGCAGATCACCTGTGGCAGCGGCCCTTGGTGAAGTGAAGGTG 420

421 GCGCGCCAGATCAAGGAGGCCCCCTGTGGACACCGCGCGAGCAGACCCGTGTGGAGGAG 480  
421 GCGCGCCAGATCAAGGAGGCCCCCTGTGGACACCGCGCGAGCAGACCCGTGTGGAGGAG 480

481 ATGAGCTGCGCGCAAGTGGAGGCCCAAGATGATCGGCGGATCGCGGCTTCAATCAAG 540  
481 ATGAGCTGCGCGCAAGTGGAGGCCCAAGATGATCGGCGGATCGCGGCTTCAATCAAG 540



```
Db 1615 GTGTACTACGACCCAGCAAGGAGCTGGTGCCGAGATCCAGAAAGCAGGGCCACGACGAG 1674
Qy 1681 TGGACCTTACCAGATCTTACCAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db 1675 TGGACCTTACCAGATCTTACCAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
Qy 1741 ATGGCCACCGCCCAACAACAACGAGTGAAGCAGCTGACCGAGCGCGTGTGCAAGAGATCGCC 1800
Db 1735 ATGGCCACCGCCCAACAACAACGAGTGAAGCAGCTGACCGAGCGCGTGTGCAAGAGATCGCC 1794
Qy 1801 ATGGAGAGCATCGTGAATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAG 1860
Db 1795 ATGGAGAGCATCGTGAATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAG 1854
Qy 1861 ACCTGGGAGACCTTGTGTGACCGGACTACTTGGCAGGCCACCTGGATCCCGAGTGGGAGTTC 1920
Db 1855 ACCTGGGAGACCTTGTGTGACCGGACTACTTGGCAGGCCACCTGGATCCCGAGTGGGAGTTC 1914
Qy 1921 GTGAACACCCCCCTGTGTGAAGCTGTGTGTAACAGCTGGAGAAAGAGGCCATCATCGGC 1980
Db 1915 GTGAACACCCCCCTGTGTGAAGCTGTGTGTAACAGCTGGAGAAAGAGGCCATCATCGGC 1974
Qy 1981 GCCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCAGACCAAGATCGGCAGGCCGCGC 2040
Db 1975 GCCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCAGACCAAGATCGGCAGGCCGCGC 2034
Qy 2041 TAGCTGACCGACCGGGCCCGGCAAGAGATCGTAGCTGTGACCGAGACCAACCAAGAG 2100
Db 2035 TAGCTGACCGACCGGGCCCGGCAAGAGATCGTAGCTGTGACCGAGACCAACCAAGAG 2094
Qy 2101 ACCGAGCTGACGCGCATCTCAGCTGGCCCTGACGACAGCGGCGAGCGGTGAACATCGTG 2160
Db 2095 ACCGAGCTGACGCGCATCTCAGCTGGCCCTGACGACAGCGGCGAGCGGTGAACATCGTG 2154
Qy 2161 ACCGACAGCAGTACGCGCTGGGATCATTCAGGCCCGAGCCCGACAGAGCGAGCGAG 2220
Db 2155 ACCGACAGCAGTACGCGCTGGGATCATTCAGGCCCGAGCCCGACAGAGCGAGCGAG 2214
Qy 2221 CTGTGTAAACAGATCATCTGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2280
Db 2215 CTGTGTAAACAGATCATCTGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2274
Qy 2281 CCCGCCCAAGGCGCATCGCGGCAACAGCAGATCGACAGCTGTGTGAGCAAGGGCATC 2340
Db 2275 CCCGCCCAAGGCGCATCGCGGCAACAGCAGATCGACAGCTGTGTGAGCAAGGGCATC 2334
Qy 2341 CGCAAGGTGCTGTCTTGGAGCGGCATCGATGGCGGCATCGTATCTACAGTACATGGAC 2400
Db 2335 CGCAAGGTGCTGTCTTGGAGCGGCATCGATGGCGGCATCGTATCTACAGTACATGGAC 2394
Qy 2401 GACTGTACCTGGGACGCGCGCTTAGGATCGATTAAAGCTTCCCGGGGCTTAGCACC 2460
Db 2395 GACTGTACCTGGGACGCGCGCTTAGGATCGATTAAAGCTTCCCGGGGCTTAGCACC 2454
Qy 2461 GGTGAATTC 2469
Db 2455 GGTGAATTC 2463
```

RESULT 6  
ADM73765

ID ADM73765 standard; DNA; 2463 BP.

XX AC ADM73765;

XX 03-JUN-2004 (first entry)

DT HIV-1 polynucleotide #8.

XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
XX HIV type C protein; immunostimulant.

Human immunodeficiency virus 1.

US2003223961-A1.

PD 04-DEC-2003.

XX 05-JUL-2001; 2001US-00899575.

PR 05-JUL-2000; 2000US-00610313.

XX (MEGE/) MEGEDE J Z.

PA (BARN/) BARNETT S W.

PA (ENGE/) ENGELBRECHT S.

PA (RENS/) RENSBUURG E J V.

XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;

XX WPI; 2004-060515/06.

XX New expression cassette comprising a polynucleotide sequence encoding an  
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
PT immunization, generating of packaging cell lines or in producing HIV Type  
PT C proteins.

PS Claim 1; SEQ ID NO 31; 160pp; English.

XX The invention relates to an expression cassette comprising a  
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
CC also relates to a recombinant expression system for use in a host cell  
CC comprising an expression cassette, where the polynucleotide sequence  
CC further comprises control elements capable of driving expression in the  
CC selected host cell, a cell comprising an expression cassette where the  
CC polynucleotide sequence further comprises control elements compatible  
CC with the expression in the cell and a composition for generating an  
CC immunological response, comprising an expression cassette. The expression  
CC cassette and the methods of the invention are useful in eliciting an  
CC immune response, in DNA immunisation, in generation of packaging cell  
CC lines and in producing HIV Type C proteins. This sequence represents an  
CC HIV-1 polynucleotide of the invention.

SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 98.9%; Score 2442.2; DB 12; Length 2463;

Best Local Similarity 99.6%; Pred. No. 1.6e-294;

Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```
Qy 1 GTGACGCCACCATGCGGAGGCCATGAGCCAGCCAGCCAGCCACCATCTCATCGAG 60
Db 1 GTGACGCCACCATGCGGAGGCCATGAGCCAGCCAGCCAGCCACCATCTCATCGAG 60
Qy 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTCGGCAAGAGGGC 120
Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTCGGCAAGAGGGC 120
Qy 121 CATATCGCCGCAACTGCGCGGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAG 180
Db 121 CATATCGCCGCAACTGCGCGGCCCGCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAG 180
Qy 181 GGCACACAGATGAAGGACTGCAACCGAGCGCCAGCCCAACTTCTCCGCGAGGACTCGCC 240
Db 181 GGCACACAGATGAAGGACTGCAACCGAGCGCCAGCCCAACTTCTCCGCGAGGACTCGCC 240
Qy 241 TTCCCCCAAGGCAAGGCCCGCGAGTTCCCGCAGGAGCAGAACCGCGCAACAGCCCCCACC 300
Db 241 TTCCCCCAAGGCAAGGCCCGCGAGTTCCCGCAGGAGCAGAACCGCGCAACAGCCCCCACC 300
Qy 301 AGCCGCGAGTGTGAGTGTGCGGCGCAACAACCCCGCAGCGAGCGCGCGCGCGAGCGCAG 360
Db 301 AGCCGCGAGTGTGAGTGTGCGGCGCAACAACCCCGCAGCGAGCGCGCGCGCGAGCGCAG 360
Qy 361 GGCACCTGTAACTTCCCGCAGATCATCCCTGTGCGAGCGGCCCTCTGTGAGCATCAAGTG 420
Db 361 GGCACCTGTAACTTCCCGCAGATCATCCCTGTGCGAGCGGCCCTCTGTGAGCATCAAGTG 420
```

QY	421	GGCGGCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCGACGACACCGTGTCTGGAGGAG	480
Db	421		
QY	481	ATGAGCCTGCCCGGCAAGTGGAAAGCCAAAGATGATCGGCGGCATCGCGGCTTTCATCAAG	540
Db	481		
QY	541	GTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAGAGAGCGCATCGGCACCGTG	600
Db	541		
QY	601	CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601		
QY	661	ACCCTGAACCTTCCCATCAAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG	720
Db	661		
QY	721	GACGCCCCAAGTGAAGATGTGGCCCCCTGACCGAGAGAGATCAAGGCCCTGACCGCC	780
Db	721		
QY	781	ATCTGCGAGAGATGAGAGAGGAGGCAAGATCACCAAGATCGCCCCCGAGAACCCCTTAC	840
Db	781		
QY	841	AACACCCCGTGTTCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGGTGGAC	900
Db	841		
QY	901	TTCCGCGAGCTGAACAGGCGACCCAGGACTTCTGGGAGGTGCGAGCTGGGCATCCCCAC	960
Db	901		
QY	961	CCCGCGGCTGAAGAAAGAGAGCGTGCACCGTGTGACGTGGCGACCGCTACTTC	1020
Db	961		
QY	1021	AGCGTCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTACCATCCCCAGCATCAAC	1080
Db	1021		
QY	1081	AACGAGACCCCGGCATTCGCTACGAGTACAACTGTCGCGCCACAGGCTGGAGGGCAGC	1140
Db	1081		
QY	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCCGCGCCCGCAACCCC	1200
Db	1141		
QY	1201	GAGATCGTGATCTACCAAGTACATGGACGACCTGTACGTGGGCGAGCGACCTGGAGATCGGC	1260
Db	1201		
QY	1261	CAGCACCGCGCCAAAGATCGAGGAGCTGCGCAAGCATCTGCTGCGCTGGGGCTTCAACACC	1320
Db	1261		
QY	1321	CCCGACAAGAACGACCAAGAGGAGCCCTTCTCTGTGATGGCTACGAGCTGCACCCC	1380
Db	1321		
QY	1381	GACAAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGAC	1440
Db	1381		
QY	1441	ATCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1500
Db	1441		
QY	1501	CGCCAGCTGTCAAGCTGTGCGCGCGCCAAAGGCCCTGACCGACATCGTGCCCTTGACC	1560
Db	1501		
QY	1561	GAGGAGCCGAGCTGGAGCTGGCCGAGAACCCGAGATCCTGGGAGCCCGTGCAACGC	1620
Db	1561		
QY	1621	GTGTACTACGACCCCAAGCAAGCACTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAG	1680
Db	1621		
QY	1681	TGACCTTACAGATCTTACAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCCAAG	1740
Db	1681		
QY	1741	ATGCGCACCGCCCAACCAACGATGAAGAGCTGACCGAGGCCGTGACAGAGATCGCC	1800
Db	1741		
QY	1801	ATGGAGAGCATCGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAG	1860
Db	1801		
QY	1861	ACCTGGGAGACCTGTGTGACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGGAGTTC	1920
Db	1861		
QY	1921	GTGAAACACCCCGCTGTGTGAGCTGTGTGATCCAGCTGGAGAGAGCCCATCATCGGC	1980
Db	1921		
QY	1981	GCGGAGACCTTCTACGTGAGACCGCGCGCAACCGCGAGACCAAGATCGCAAGGCGCGGC	2040
Db	1981		
QY	2041	TACGTGACCGACCGGGCGCGGAGAAAGATCGTGAGCCTGACCGAGACCAACCAAGAAAG	2100
Db	2041		
QY	2101	ACCGAGCTCAGGCCCATCCAGCTGGCCCTGCAAGGACAGCGGCGAGGCTGAACATCGTG	2160
Db	2101		
QY	2161	ACCGAGCTCAGGCCCATCCAGCTGGCCCTGCAAGGACAGCGGCGAGGCTGAACATCGTG	2220
Db	2161		
QY	2221	CTGGTGAACCAAGATCATCGAGCAGCTGTATCAAGAGGAGAAAGGTGTACTGTAGCTGGGTG	2280
Db	2221		
QY	2281	CCGCCCCACAGGGCATCGGCGCAACGAGCAGATCGACAGCTGGTGGTGAAGGGGCATC	2340
Db	2281		
QY	2341	CGCAAGTGTCTTCTCTGGACCGCATCGATGGCGCATCGTGATCTACCAAGTACATGGAC	2400
Db	2341		
QY	2401	GACCTGTAGTGGGCGAGCGGCCCTTAGATCGATTAAAGTTCCCGGGGCTAGCACC	2460
Db	2401		
QY	2461	GGTGAATTC	2469
Db	2461		
QY	2469	GGTGAATTC	2475
Db	2469		

RESULT 7

ACA03547

ID ACA03547 standard; DNA; 2457 BP.

XX



Db 1315 CCCGACAAGAAGCACACAGAGGAGCCCCCTTCTCTGTGGATGGCTACGAGCTGCACCCC 1374

QY 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCCGTGAAGAC 1440

Db 1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCCGTGAAGAC 1434

QY 1441 ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG 1500

Db 1435 ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG 1494

QY 1501 CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACGACATCTGTGCCCTGACC 1560

Db 1495 CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACGACATCTGTGCCCTGACC 1554

QY 1561 GAGGAGCCGAGCTGAGCTGTGGCCGAGAACCCGCGAGATCTGCGCGAGCCCGTGCAGGC 1620

Db 1555 GAGGAGCCGAGCTGAGCTGTGGCCGAGAACCCGCGAGATCTGCGCGAGCCCGTGCAGGC 1614

QY 1621 GTCTACTACGACCCCAAGAGGACCTGTGGCGGAGATCCAGAAGCAGGCGCACGACCAG 1680

Db 1615 GTCTACTACGACCCCAAGAGGACCTGTGGCGGAGATCCAGAAGCAGGCGCACGACCAG 1674

QY 1681 TGGACCTTACAGATCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1740

Db 1675 TGGACCTTACAGATCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1734

QY 1741 ATCGCACCGCCACACACAGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAGATCGCC 1800

Db 1735 ATCGCACCGCCACACACAGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAGATCGCC 1794

QY 1801 ATCGAGAGCATCTGTGCTGGGCGAGACCCCAAGTTCCGCTGCCCATCCAGAAGGAG 1860

Db 1795 ATCGAGAGCATCTGTGCTGGGCGAGACCCCAAGTTCCGCTGCCCATCCAGAAGGAG 1854

QY 1861 ACCTGGGAGACCTGTGGACCGACTACTGCGAGGCCACCTGGATCCCCCGAGTGGAGTTC 1920

Db 1855 ACCTGGGAGACCTGTGGACCGACTACTGCGAGGCCACCTGGATCCCCCGAGTGGAGTTC 1914

QY 1921 GTGAACACCCCCCTGTGTGAAGCTGTGTGTACAGCTGAGAGAGGCCCATCATCGGC 1980

Db 1915 GTGAACACCCCCCTGTGTGAAGCTGTGTGTACAGCTGAGAGAGGCCCATCATCGGC 1974

QY 1981 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCG 2040

Db 1975 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGCG 2034

QY 2041 TACGTGACCGAGCGGGCCCGGAGAGATCGTGAGCTTGACCGAGACCAACACCAAG 2100

Db 2035 TACGTGACCGAGCGGGCCCGGAGAGATCGTGAGCTTGACCGAGACCAACACCAAG 2094

QY 2101 ACCGAGCTGAGGCGCATTCAGCTGGCCCTGCAGGACGCGGCGAGCGGTGAACATCGTG 2160

Db 2095 ACCGAGCTGAGGCGCATTCAGCTGGCCCTGCAGGACGCGGCGAGCGGTGAACATCGTG 2154

QY 2161 ACCGAGCGCAGTACGCGCTTGGGCATCATCCAGGCCCGAGCCGACAGAGCGAGCGAG 2220

Db 2155 ACCGAGCGCAGTACGCGCTTGGGCATCATCCAGGCCCGAGCCGACAGAGCGAGCGAG 2214

QY 2221 CTGCTGAACACAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTCTGAGCTGGTG 2280

Db 2215 CTGCTGAACACAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTCTGAGCTGGTG 2274

QY 2281 CCCGCCACAAAGGCGATCGCGGCGACGAGCAGATCGACAAGCTGGTGAGCAAGGCGATC 2340

Db 2275 CCCGCCACAAAGGCGATCGCGGCGACGAGCAGATCGACAAGCTGGTGAGCAAGGCGATC 2334

QY 2341 CGCAAGTGTCTGTCTTGGAGCGCATCGATGGCGGCATCTGTATCTACCAAGTACATGGAC 2400

Db 2335 CGCAAGTGTCTGTCTTGGAGCGCATCGATGGCGGCATCTGTATCTACCAAGTACATGGAC 2394

QY 2401 GACCTGTACGTGGGCGAGCGCGCCCTTAGGATCGATTAAGAGCTTCCCGGGCTTAGCACC 2460

Db 2395 GACCTGTACGTGGGCGAGCGCGCCCTTAGGATCGATTAAGAGCTTCCCGGGCTTAGCACC 2454

QY 2461 GGT 2463

Db 2455 GGT 2457

RESULT 8

ADCL3265

ID ADCL3265 standard; DNA; 2457 BP.

XX

AC ADCL3265;

DT 18-DEC-2003 (first entry)

XX

DE DNA of HIV construct p2Pol-opt-YM\_C SEQ ID NO 44.

XX

KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX

OS Human immunodeficiency virus.

XX

PN WO2003004620-A2.

XX

PD 16-JAN-2003.

XX

PF 05-JUL-2002; 2002WO-US021420.

XX

PR 05-JUL-2001; 2001US-0303192P.

PR 31-AUG-2001; 2001US-0316860P.

PR 16-JAN-2002; 2002US-0349871P.

XX

PA (CHIR ) CHIRON CORP.

PA (UYST-) UNIV STELLENBOSCH.

XX

Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

WPI; 2003-221593/21.

New expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide, useful for immunization, or generating packaging cell lines.

Disclosure; Fig 41; 301pp; English.

The invention relates to a novel expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel expression cassette can be used to treat HIV type C by gene therapy or used in the development of a vaccine. The gene delivery vector is administered intramuscularly, intramucosally, intranasally, subcutaneously, intradermally, transdermally, intravaginally, CC intrateally, orally or intravenously. The expression cassette is useful for immunisation, generating packaging cell lines and producing HIV polypeptides. This polynucleotide sequence represents the DNA of an HIV Type C related sequence of the invention.

Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.6%; Score 2434.6; DB 10; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 1.4e-293;  
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 1 GTGAGCGCCACCATTGCGCGGCGCATGAGCCAGCCAGCCACCATCTCTGATGCAG 60

Db 1 GTGAGCGCCACCATTGCGCGGCGCATGAGCCAGCCAGCCACCATCTCTGATGCAG 60

QY 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC 120

Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGC 120

QY 121 CACATCGCCGCAACTGCGCGCCCCCGGCAAGAGGGTGTCTGGAAGTCCGGCAAGAG 180

Db 121 CACATCGCCGCACTCGCGCCGCCCCCGCAAGAGGCTGCTGGAAGTGGGCAAGGAG 180  
Qy 181 GGCCACAGATGAAGGACTGACAGCGCCGCAAGGCAACTTTCTTCGCGAGGAACTTGGCC 240  
Db 181 GGCCACAGATGAAGGACTGACAGCGCCGCAAGGCAACTTTCTTCGCGAGGAACTTGGCC 240  
Qy 241 TTCCCCAGGCGCAAGGCGCGAGTTCCCGAGCGAGCAACCCGCGCCAAAGCCGCGCC 300  
Db 241 TTCCCCAGGCGCAAGGCGCGAGTTCCCGAGCGAGCAACCCGCGCCAAAGCCGCGCC 300  
Qy 301 AGCCGCGAGCTGCAAGTGGCGGCAAAACCCGCGAGCGCGCGCGCGCGCGCGCGCG 360  
Db 301 AGCCGCGAGCTGCAAGTGGCGGCAAAACCCGCGAGCGCGCGCGCGCGCGCGCGCG 360  
Qy 361 GGCACTCTGAATCTTCCCGCAGATCACTCTGTGGCAGCGCCCTCTGGTGAGCATCAAGTG 420  
Db 361 GGCACTCTGAATCTTCCCGCAGATCACTCTGTGGCAGCGCCCTCTGGTGAGCATCAAGTG 420  
Qy 421 GGCGGCGAGATCAAGGAGGCTCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Db 421 GGCGGCGAGATCAAGGAGGCTCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCG 480  
Qy 481 ATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAACAG 540  
Db 481 ATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAACAG 540  
Qy 541 GTGCGCCAGTACGACCAAGTCTGATCGAGATCTCGCGCAAGAGGCGCATCGGCAACG 600  
Db 541 GTGCGCCAGTACGACCAAGTCTGATCGAGATCTCGCGCAAGAGGCGCATCGGCAACG 600  
Qy 601 CTGATCGGCGCCACCGCGTGAAATCATCTCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
Db 601 CTGATCGGCGCCACCGCGTGAAATCATCTCGCGCGCAACATGCTGACCCAGCTGGGTGC 660  
Qy 661 ACCGTGAATCTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGTGAAGCGCGCATG 720  
Db 661 ACCGTGAATCTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGTGAAGCGCGCATG 720  
Qy 721 GACGCGCCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 780  
Db 721 GACGCGCCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 780  
Qy 781 ATCTGCGAGGAGATGGAAGAGGCGGCAAGATCAACAGATCGGCGCGCGCAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAAGAGGCGGCAAGATCAACAGATCGGCGCGCGCAACCCCTAC 840  
Qy 841 AACACCCCGTGTTCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
Db 841 AACACCCCGTGTTCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
Qy 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGCAC 960  
Db 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGCAC 960  
Qy 961 CCCGCGGCTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
Db 961 CCCGCGGCTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
Qy 1021 AGCGTGCCCTGGAGAGGAGGAGTTCGCAAGTGAACCGCTTCAACATCCCGCAGCATCAAC 1080  
Db 1021 AGCGTGCCCTGGAGAGGAGGAGTTCGCAAGTGAACCGCTTCAACATCCCGCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGCATCCGCTACAGTACAACTGTGCTGCCCGCGAGGCTGGAGGCGAGC 1140  
Db 1081 AACGAGACCCCGCGCATCCGCTACAGTACAACTGTGCTGCCCGCGAGGCTGGAGGCGAGC 1140  
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGGCGCTTCCGCGCGCGCAACCCC 1200  
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGGCGCTTCCGCGCGCGCAACCCC 1200  
Qy 1201 GAGATCGATCTACAGTACATGAGCGACTGTGCTGGGCGAGGAGGAGGAGGAGGAGGAG 1260  
Db 1201 GAGATCGATCTACAGTACATGAGCGACTGTGCTGGGCGAGGAGGAGGAGGAGGAGGAG 1260  
1201 GAGATCGATCTACAGTACATGAGCGACTGTGCTGGGCGAGGAGGAGGAGGAGGAGGAG 1254

Qy 1261 CAGCAACCGCGCAAGATCGAGGAGTGGCGCAAGCACTGCTGCTGGGCTTCAACCAACC 1320  
Db 1255 CAGCAACCGCGCAAGATCGAGGAGTGGCGCAAGCACTGCTGCTGGGCTTCAACCAACC 1314  
Qy 1321 CCCGACAGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
Db 1315 CCCGACAGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1374  
Qy 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAG 1440  
Db 1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAG 1434  
Qy 1441 ATCCAGAGCTGTGTGGGCAAGTGAATGGGCGCAGGCGAGATCTACCCCGGCGATCAAGGTG 1500  
Db 1435 ATCCAGAGCTGTGTGGGCAAGTGAATGGGCGCAGGAGATCTACCCCGGCGATCAAGGTG 1494  
Qy 1501 CGCCAGCTGTGCAAGCTGTGCGGCGGCGCAAGGCGCTGACCGGAGATCGTGCCCTGAGC 1560  
Db 1495 CGCCAGCTGTGCAAGCTGTGCGGCGGCGCAAGGCGCTGACCGGAGATCGTGCCCTGAGC 1554  
Qy 1561 GAGGAGGCGGAGCTGTGAGCTGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGACGCGC 1620  
Db 1555 GAGGAGGCGGAGCTGTGAGCTGCGCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGACGCGC 1614  
Qy 1621 GTGTACTACGACCCAGCAAGGACTGTGTGGCGCGAGATCTCAGAAAGAGGCGCAACGACGAG 1680  
Db 1615 GTGTACTACGACCCAGCAAGGACTGTGTGGCGCGAGATCTCAGAAAGAGGCGCAACGACGAG 1674  
Qy 1681 TGGACCTTACCAAGTCTACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
Db 1675 TGGACCTTACCAAGTCTACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1734  
Qy 1741 ATGCGCACCGGCGCACCAAGAGTGAAGCAGCTGACCGAGGCGCTGACAGAGATCGCC 1800  
Db 1735 ATGCGCACCGGCGCACCAAGAGTGAAGCAGCTGACCGAGGCGCTGACAGAGATCGCC 1794  
Qy 1801 ATGAGAGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGGAG 1860  
Db 1795 ATGAGAGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGGAG 1854  
Qy 1861 ACCTGGGAGAGCTTGGTGACCGACTACTGCGAGCGCCACCTGGATCCCGAGTGGGAGTTC 1920  
Db 1855 ACCTGGGAGAGCTTGGTGACCGACTACTGCGAGCGCCACCTGGATCCCGAGTGGGAGTTC 1914  
Qy 1921 GTGAACACCCCGGCGCGGAGCTGTGTGATGAGCTGTGTGAGTGTGTGAGAGGAGGAGGAGGAG 1980  
Db 1915 GTGAACACCCCGGCGCGGAGCTGTGTGATGAGCTGTGTGAGTGTGTGAGAGGAGGAGGAGGAG 1974  
Qy 1981 GCCGAGACCTTTCTACGTGGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCGC 2040  
Db 1975 GCCGAGACCTTTCTACGTGGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCGC 2034  
Qy 2041 TAGCTGACCGACCGGCGCGGAGAGATCGTGAGCTTACCGAGACCGACCAACCGAGAG 2100  
Db 2035 TAGCTGACCGACCGGCGCGGAGAGATCGTGAGCTTACCGAGACCGACCAACCGAGAG 2094  
Qy 2101 ACCGAGCTGACGAGGCGCATCAGCTGCGCGCTGCGAGCAGCGGAGGAGGAGGAGGAGGAGGAG 2160  
Db 2095 ACCGAGCTGACGAGGCGCATCAGCTGCGCGCTGCGAGCAGCGGAGGAGGAGGAGGAGGAGGAG 2154  
Qy 2161 ACCGAGCTGACGAGGCGCATCAGCTGCGCGCTGCGAGCAGCGGAGGAGGAGGAGGAGGAGGAG 2220  
Db 2155 ACCGAGCTGACGAGGCGCATCAGCTGCGCGCTGCGAGCAGCGGAGGAGGAGGAGGAGGAGGAG 2214  
Qy 2221 CTGTGAAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280  
Db 2215 CTGTGAAACAGATCATCGAGCAGCTGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2274  
Qy 2281 CCCGCGCACAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGTGAGCAAGGAGGAGGAGGAGGAG 2340  
Db 2275 CCCGCGCACAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGTGAGCAAGGAGGAGGAGGAGGAG 2334

QY	2341	CGCAAGTGCTGTTCTCTGGACGCGCATCGATGCGCGGCATCGTGATCTACCAAGTACATGGAC	2400
DB	2335	CGCAAGTGCTGTTCTCTGGACGCGCATCGATGCGCGGCATCGTGATCTACCAAGTACATGGAC	2394
QY	2401	GACCTGTACGTGGCGACGCGCGCCCTAGGATCGATTTAAAGTCTCCCGGGCTAGCACC	2460
DB	2395	GACCTGTACGTGGCGACGCGCGCCCTAGGATCGATTTAAAGTCTCCCGGGCTAGCACC	2454
QY	2461	GGT 2463	
DB	2455	GGT 2457	
RESULT 9			
ID	ABL39961	standard; DNA; 2457 BP.	
AC	ABL39961;		
XX	15-MAY-2002	(first entry)	
XX	Synthetic construct PR975YMM SEQ ID NO:32.		
DE	Human immunodeficiency virus type C; antigenic HIV type C protein;		
XX	immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;		
KW	immunostimulant; gene therapy; gene; ds.		
XX	Human immunodeficiency virus; type C.		
OS	Synthetic.		
XX	WO200204493-A2.		
PN	17-JAN-2002.		
XX	05-JUL-2001; 2001WO-US021241.		
XX	05-JUL-2000; 2000US-00610313.		
PR	(CHIR ) CHIRON CORP.		
XX	(UYST-) UNIV STELLENBOSCH.		
PA	Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;		
XX	WPI; 2002-154920/20.		
DR	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in		
XX	applications including DNA immunization or generation of packaging cell		
PT	lines, particularly in gene therapy.		
XX	Claim 1; Fig 10; 233pp; English.		
PS	The present invention describes expression cassettes comprising a		
CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV		
CC	type C polypeptides. The expression cassettes comprise any of the HIV		
CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef		
CC	(I). (I) have immunostimulant activity and can be used in gene therapy.		
CC	The HIV type C polynucleotides are useful in applications including DNA		
CC	immunisation, generation of packaging cell lines, and production of HIV		
CC	type C proteins. The polynucleotides are particularly useful in gene		
CC	therapy and DNA immunisation applications. ABL39942 to ABL40054 and		
CC	ABB06204 to ABB06215 represent sequences used in the exemplification of		
CC	the present invention		
XX	Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;		
SQ	Query Match 97.8%; Score 2415.4; DB 6; Length 2457;		
	Best Local Similarity 99.3%; Pred. No. 3.3e-291;		
	Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;		
QY	1	GTGACGCCACCATGGCCGAGGCGCATGAGCCAGGCCACCGCCCAACATCTCTGATGCAG	60
DB	1	GTGACGCCACCATGGCCGAGGCGCATGAGCCAGGCCACCGCCCAACATCTCTGATGCAG	60

QY	61	CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC	120
DB	61	CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC	120
QY	121	CACATCGCGCGCAACTGCGCGCCCCCGCAAGAGGGTGTCTGGAAGTGCGCAGAGGAG	180
DB	121	CACATCGCGCGCAACTGCGCGCCCCCGCAAGAGGGTGTCTGGAAGTGCGCAGAGGAG	180
QY	181	GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGGAGGACCTTGCC	240
DB	181	GGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGGAGGACCTTGCC	240
QY	241	TTCCCCCAGGCGAAGGCCCGCGAGTTTCCCCAGCGAGCAGAACCGGCCCAACAGCCCCACC	300
DB	241	TTCCCCCAGGCGAAGGCCCGCGAGTTTCCCCAGCGAGCAGAACCGGCCCAACAGCCCCACC	300
QY	301	AGCCGCGAGCTGCAAGTGTGCGCGCAAAACCCCGCAGCGAGGCCCGCGCGAGCGCAG	360
DB	301	AGCCGCGAGCTGCAAGTGTGCGCGCAAAACCCCGCAGCGAGGCCCGCGCGAGCGCAG	360
QY	361	GGCACCCTGAATTTCCCCCAGATCAACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG	420
DB	361	GGCACCCTGAATTTCCCCCAGATCAACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG	420
QY	421	GGCGGCGAGATCAAGGAGGCGCTGTCTGGACACCGCGCGCGACGACACCGTGTGGAGGAG	480
DB	421	GGCGGCGAGATCAAGGAGGCGCTGTCTGGACACCGCGCGCGACGACACCGTGTGGAGGAG	480
QY	481	ATGAGCCTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG	540
DB	481	ATGAGCCTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG	540
QY	541	GTGCGCGAGTACGACCGAGATCTGTGTCGAGATCTGCGGCAAGAGGCCATCGCACCGTG	600
DB	541	GTGCGCGAGTACGACCGAGATCTGTGTCGAGATCTGCGGCAAGAGGCCATCGCACCGTG	600
QY	601	CTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC	660
DB	601	CTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC	660
QY	661	ACCTGAACTTCCCATCGACCCCATCGAGACCGTCCCGTGAAGTGAAGCCCGGCGATG	720
DB	661	ACCTGAACTTCCCATCGACCCCATCGAGACCGTCCCGTGAAGTGAAGCCCGGCGATG	720
QY	721	GACGCGCCCAAGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCCCTGACCGCC	780
DB	721	GACGCGCCCAAGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCCCTGACCGCC	780
QY	781	ATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGAAACCCCTAC	840
DB	781	ATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGAAACCCCTAC	840
QY	841	AACACCCCGTGTTCGCATCAAGAAAGAGGACAGCACCAAGTGGGCGCAAGCTGGTGAC	900
DB	841	AACACCCCGTGTTCGCATCAAGAAAGAGGACAGCACCAAGTGGGCGCAAGCTGGTGAC	900
QY	901	TTTCGCGAGCTGCAACCAAGCGCACCCAGGACTTCTGCGAGGTGCGAGTGGGCATCCCCAC	960
DB	901	TTTCGCGAGCTGCAACCAAGCGCACCCAGGACTTCTGCGAGGTGCGAGTGGGCATCCCCAC	960
QY	961	CCCGCGGCTGAAAGAAAGAGAGGAGCGTGACCGTGTCTGACGTGGGCGACGCTACTTC	1020
DB	961	CCCGCGGCTGAAAGAAAGAGAGGAGCGTGACCGTGTCTGACGTGGGCGACGCTACTTC	1020
QY	1021	AGCGTGCCTTGACCGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC	1080
DB	1021	AGCGTGCCTTGACCGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC	1080
QY	1081	AACGACACCCCGGCGATCCGCTACCAAGTACAACGTCTGCCCGCAGGGCTGGAAGGGCAGC	1140
DB	1081	AACGACACCCCGGCGATCCGCTACCAAGTACAACGTCTGCCCGCAGGGCTGGAAGGGCAGC	1140
QY	1141	CCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTTCGCGCGCCGCAACCCC	1200





Query Match		97.8%	Score 2415, 4:	DB 12:	Length 2457;
Best Local Similarity		99.3%	Pred. No. 3.3e-291;		
Matches 2451;		Conservative 0;	Mismatches 6;	Indels 12;	Gaps 2;
Qy	1	GTGAGGCCACCATGGCCGAGGCGCATGAGCAGGCGCCACACAGCGCCCAACATCTTGATGCGAG	60		
Db	1	GTGAGGCCACCATGGCCGAGGCGCATGAGCAGGCGCCACACAGCGCCCAACATCTTGATGCGAG	60		
Qy	61	CGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTTCATCACTGCGGCAAGAGGGG	120		
Db	61	CGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTTCATCACTGCGGCAAGAGGGG	120		
Qy	121	CACATGCGCGCAACTGCGCGCCCCCGCAGCAAGAGGGCTGCTGGAAGTGGCGCAAGAG	180		
Db	121	CACATGCGCGCAACTGCGCGCCCCCGCAGCAAGAGGGCTGCTGGAAGTGGCGCAAGAG	180		
Qy	181	GGCACACAGATGAAGGACTGCACCGAGCGCCAGCGCCAACTTCTTCGCGGAGGACCTGGCC	240		
Db	181	GGCACACAGATGAAGGACTGCACCGAGCGCCAGCGCCAACTTCTTCGCGGAGGACCTGGCC	240		
Qy	241	TTCCCCAGGGCAAGGCGCGGAGTTCCCGCAGCGAGCAGAAACCGCGGCCAACAGCCCCACC	300		
Db	241	TTCCCCAGGGCAAGGCGCGGAGTTCCCGCAGCGAGCAGAAACCGCGGCCAACAGCCCCACC	300		
Qy	301	AGCGCGAGCTGCGAGTGGCGGCGCAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAG	360		
Db	301	AGCGCGAGCTGCGAGTGGCGGCGCAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAG	360		
Qy	361	GGCACCTGAACTTCCCCCAGATCACTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG	420		
Db	361	GGCACCTGAACTTCCCCCAGATCACTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG	420		
Qy	421	GGCGGCGAGTCAAGAGGGCCCTGTGGACACCGCGCGCGCAGACACCGTGTGGAGGAG	480		
Db	421	GGCGGCGAGTCAAGAGGGCCCTGTGGACACCGCGCGCGCAGACACCGTGTGGAGGAG	480		
Qy	481	ATGAGCTGCGCGCAAGTGGAGGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAG	540		
Db	481	ATGAGCTGCGCGCAAGTGGAGGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAG	540		
Qy	541	GTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCATCGGCACCGTG	600		
Db	541	GTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCATCGGCACCGTG	600		
Qy	601	CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCGAGCTGGGCTGC	660		
Db	601	CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCGAGCTGGGCTGC	660		
Qy	661	ACCTTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720		
Db	661	ACCTTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720		
Qy	721	GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780		
Db	721	GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780		
Qy	781	ATCTGGAGAGATGAGAGAGGAGGCGCAAGATCACCAAGATCGGCGCCCGAGAACCCCTTAC	840		
Db	781	ATCTGGAGAGATGAGAGAGGAGGCGCAAGATCACCAAGATCGGCGCCCGAGAACCCCTTAC	840		
Qy	841	AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900		
Db	841	AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900		
Qy	901	TTCCGGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCAC	960		
Db	901	TTCCGGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCAC	960		
Qy	961	CCCGCGGCTGGAAGAGAGAGCGGTGACCGTGTGAGAGTGGCGAGCGCTTACTTC	1020		
Db	961	CCCGCGGCTGGAAGAGAGAGCGGTGACCGTGTGAGAGTGGCGAGCGCTTACTTC	1020		

Qy	1021	AGCGTCCCTTGACGAGGACTTCCGAAAGTACACCGCCTTACCATTCCCGAGCATCAAC	1080		
Db	1021	AGCGTCCCTTGACGAGGACTTCCGAAAGTACACCGCCTTACCATTCCCGAGCATCAAC	1080		
Qy	1081	AACGAGACCCCGGCATCCGCTACAGTACAAGTCTGCTGCCCCAGGGCTGGAAAGGGCAGC	1140		
Db	1081	AACGAGACCCCGGCATCCGCTACAGTACAAGTCTGCTGCCCCAGGGCTGGAAAGGGCAGC	1140		
Qy	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC	1200		
Db	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCC	1200		
Qy	1201	GAGATCGTATCTTACAGTACATGAGCAGCTGTAGTGGGAGCGACCTTGGAGATCGGC	1260		
Db	1201	GAGATCGTATCTTACCA-----GGCCCCCTGTAGTGGGAGCGACCTTGGAGATCGGC	1254		
Qy	1261	CAGCACCGGCCAAGATCGAGGAGCTGCGCAAGCACTGTGCTGCTGGGGCTTTCACACC	1320		
Db	1255	CAGCACCGGCCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCTGCTGGGGCTTTCACACC	1314		
Qy	1321	CCCGACAAGACCAAGAGGAGCGCCCTTCTGTGTGATGGCTACGAGCTGCACCCC	1380		
Db	1315	CCCGACAAGAGCACAGAGGAGCGCCCTTCTGTGTGATGGCTACGAGCTGCACCCC	1368		
Qy	1381	GACAGTGGACCGCTGAGCCCATCGAGCTGCCCCGAGAAGGAGAGCTGGACCGTGAACGAC	1440		
Db	1369	GACAGTGGACCGCTGAGCCCATCGAGCTGCCCCGAGAAGGAGAGCTGGACCGTGAACGAC	1428		
Qy	1441	ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1500		
Db	1429	ATCCAGAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1488		
Qy	1501	CGCCAGCTGTGAAGCTGTGCGGCGCCAGCGCCCTGACCGACATCTGTCGCCCTGACCC	1560		
Db	1489	CGCCAGCTGTGAAGCTGTGCGGCGCCAGCGCCCTGACCGACATCTGTCGCCCTGACCC	1548		
Qy	1561	GAGGAGCGGAGCTGAGCTGGCCGAGAACCGCGAGATCTGTCGCGAGGCCCGTGCACGGC	1620		
Db	1549	GAGGAGCGGAGCTGAGCTGGCCGAGAACCGCGAGATCTGTCGCGAGGCCCGTGCACGGC	1608		
Qy	1621	GTGTACTACGACCCCAAGAGGACCTGTGGTGGCCGAGATTCAGAAAGCAGGCGCCACGAC	1680		
Db	1609	GTGTACTACGACCCCAAGAGGACCTGTGGTGGCCGAGATTCAGAAAGCAGGCGCCACGAC	1668		
Qy	1681	TGGACCTTACAGATCTACGAGGAGCCCTTCAAGAACTGAAGACCGCAAGTACGCCAAG	1740		
Db	1669	TGGACCTTACAGATCTACGAGGAGCCCTTCAAGAACTGAAGACCGCAAGTACGCCAAG	1728		
Qy	1741	ATCGGACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCC	1800		
Db	1729	ATCGGACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCC	1788		
Qy	1801	ATGGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCGCATTCAGAAAGGAG	1860		
Db	1789	ATGGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCGCATTCAGAAAGGAG	1848		
Qy	1861	ACCTGGAGACCTGTGTGGACCGGACTTGTGCGAGCCACCTTGGATCCCGAGTGGAGTTC	1920		
Db	1849	ACCTGGAGACCTGTGTGGACCGGACTTGTGCGAGCCACCTTGGATCCCGAGTGGAGTTC	1908		
Qy	1921	GTGAACAACCCCGCTGTGTGAAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCGGC	1980		
Db	1909	GTGAACAACCCCGCTGTGTGAAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCGGC	1968		
Qy	1981	GCGGAGACCTTCTAGCTGAGCGCGCCCAACCGGAGACCAAGATCGGAGAGCGCGGC	2040		
Db	1969	GCGGAGACCTTCTAGCTGAGCGCGCCCAACCGGAGACCAAGATCGGAGAGCGCGGC	2028		
Qy	2041	TAGCTGACCGAGCGCGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG	2100		
Db	2029	TAGCTGACCGAGCGCGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG	2088		
Qy	2101	ACCGAGCTGAGGCCATTCAGCTGGCCCTTGCAGGACAGCGGCGAGCGAGGTGAACATCGTG	2160		



Db 2089 ACCGAGCTGCAGGCGATCCAGCTGGCCCTGCAGGACAGCGGCGAGGTGAAATCATCGTG 2148  
 Qy 2161 ACCGACAGCCAGTACGGCCCTGGGCGATCATCCAGGCCCCAGCCCGGACAAAGAGCGAGCGAG 2220  
 Db 2149 ACCGACAGCCAGTACGGCCCTGGGCGATCATCCAGGCCCCAGCCCGGACAAAGAGCGAGCGAG 2208  
 Qy 2221 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAAAGGTGTACCTGAGCTGGGTG 2280  
 Db 2209 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAAAGGTGTACCTGAGCTGGGTG 2268  
 Qy 2281 CCCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAGCTGCTGAGCAAGGGGCATC 2340  
 Db 2269 CCCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAGCTGCTGAGCAAGGGGCATC 2328  
 Qy 2341 CGCAAGGTGCTGTTCTTCCGAGCGGCGATCGATCGCGGCGATCGTGTATCTACCAAGTACATGGAC 2400  
 Db 2329 CGCAAGGTGCTGTTCTTCCGAGCGGCGATCGATCGCGGCGATCGTGTATCTACCAAGTACATGGAC 2388  
 Qy 2401 GACCTGTACGTGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2460  
 Db 2389 GACCTGTACGTGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCAC 2448  
 Qy 2461 GGTGAATTC 2469  
 Db 2449 GGTGAATTC 2457

RESULT 11

ACA03546  
 ID ACA03546 standard; DNA; 2445 BP.  
 XX ACA03546;  
 AC  
 DT 22-MAY-2003 (first entry)  
 XX  
 DE Synthetic DNA encoding immunogenic HIV peptide #29.  
 XX  
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 OS Synthetic.  
 XX  
 XX WO2003004657-A1.  
 XX  
 XX 16-JAN-2003.  
 XX  
 XX 05-JUL-2002; 2002WO-US021421.  
 XX  
 XX 05-JUL-2001; 2001US-0303192P.  
 XX 31-AUG-2001; 2001US-0316860P.  
 XX 16-JAN-2002; 2002US-0349728P.  
 XX 16-JAN-2002; 2002US-0349733P.  
 XX 16-JAN-2002; 2002US-0349871P.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Zur Megede J, Barnett SW, Lian Y;  
 XX WPI; 2003-221602/21.  
 XX  
 XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX  
 XX Example 1; Fig 34; 262pp; English.  
 PS  
 XX The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV

CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 XX Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;  
 Query Match 97.3%; Score 2401.8; DB 8; Length 2445;  
 Best Local Similarity 99.2%; Pred. No. 1.6e-289;  
 Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;  
 Qy 7 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACACAGCGCCACATCTCTGATCGAGCGCAGC 66  
 Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACACAGCGCCACATCTCTGATCGAGCGCAGC 60  
 Qy 67 AACTTCAAGGGCCCAAGCGGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCGCACATC 126  
 Db 61 AACTTCAAGGGCCCAAGCGGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCGCACATC 120  
 Qy 127 GCCCGCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAAGTGGCGCAAGAGGGGCGCAC 186  
 Db 121 GCCCGCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAAGTGGCGCAAGAGGGGCGCAC 180  
 Qy 187 CAGATGAAGGACTGCG 246  
 Db 181 CAGATGAAGGACTGCG 240  
 Qy 247 CAGGGCAAGGCG 306  
 Db 241 CAGGGCAAGGCG 300  
 Qy 307 GAGCTGAGTGGCG 366  
 Db 301 GAGCTGAGTGGCG 360  
 Qy 367 CTGAATCTTCCCGCAGATCACCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426  
 Db 361 CTGAATCTTCCCGCAGATCACCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
 Qy 427 CAGATCAAGGAGGCG 486  
 Db 421 CAGATCAAGGAGGCG 480  
 Qy 487 CTGCGCGGCAAGTGGAAAGCG 546  
 Db 481 CTGCGCGGCAAGTGGAAAGCG 540  
 Qy 547 CAGTACGACGAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTGTCTGATC 606  
 Db 541 CAGTACGACGAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTGTCTGATC 600  
 Qy 607 GGCCCGCAAGGCG 666  
 Db 601 GGCCCGCAAGGCG 660  
 Qy 667 AACTTCCCGCATGAGCG 726  
 Db 661 AACTTCCCGCATGAGCG 720  
 Qy 727 CCAAGGTGAAGCAGTGGCG 786  
 Db 721 CCAAGGTGAAGCAGTGGCG 780  
 Qy 787 GAGGAGATGGAGAGGAGGCG 846  
 Db 781 GAGGAGATGGAGAGGAGGCG 840  
 Qy 847 CCGGTGTTCGCCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 906

Db 841 CCGGTGTTGCCATCAAGAAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900  
 QY 907 GAGCTGAACAGCGCACCAGGATTTCTGGAGGTTCAGAGTGGCATCCCGCAGCCCGCC 966  
 Db 901 GAGCTGAACAGCGCACCAGGATTTCTGGAGGTTCAGAGTGGCATCCCGCAGCCCGCC 960  
 QY 967 GGCCTGAAGAAGAAAGAGCGGTGACCGTCTGGAGTGGCGCAAGCTTACTTTCAGCGTG 1026  
 Db 961 GGCCTGAAGAAGAAAGAGCGGTGACCGTCTGGAGTGGCGCAAGCTTACTTTCAGCGTG 1020  
 QY 1027 CCCTTGAGACGAGACTTCCGCAAGTACACCGCTTTCACATCCCGCAGCATCAACACGAG 1086  
 Db 1021 CCCTTGAGACGAGACTTCCGCAAGTACACCGCTTTCACATCCCGCAGCATCAACACGAG 1080  
 QY 1087 ACCCCGGGATCCGCTACAGTACAGTGTCTGCGGCTGCGGCTGGAGGCGAGCCCGAC 1146  
 Db 1081 ACCCCGGGATCCGCTACAGTACAGTGTCTGCGGCTGCGGCTGGAGGCGAGCCCGAC 1140  
 QY 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
 Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200  
 QY 1207 GTGATCTACAGTACATGAGCAGCTGTGTACGTGGGAGCGACCTGTGAGATGCGCGACAC 1266  
 Db 1201 GTGATCTACCA-----GGCCCCCTGTACGTGGGAGCGACCTGTGAGATGCGCGACAC 1254  
 QY 1267 CGGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGGCTGGGGCTTCAACACCCCGAC 1326  
 Db 1255 CGGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGGCTGGGGCTTCAACACCCCGAC 1314  
 QY 1327 AAGAGCACCAGAGAGGAGCCCTTCTGTGTGTGGCTTACGAGCTGACACCCCGCAAG 1386  
 Db 1315 AAGAGCACCAGAGAGGAGCCCTTCTGTGTGTGGCTTACGAGCTGACACCCCGCAAG 1368  
 QY 1387 TGACCGGTGACGCCATCGAGCTGCCGAGAAGAGAGTGGACCGTGAACGACATCCAG 1446  
 Db 1369 TGACCGGTGACGCCATCGAGCTGCCGAGAAGAGAGTGGACCGTGAACGACATCCAG 1428  
 QY 1447 AAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGAG 1506  
 Db 1429 AAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGAG 1488  
 QY 1507 CTGTGAAGCTGTCTGCGCGCGCCCAAGGCCCTGACGACATCTGTCGCCCTGACCGAGGAG 1566  
 Db 1489 CTGTGAAGCTGTCTGCGCGCGCCCAAGGCCCTGACGACATCTGTCGCCCTGACCGAGGAG 1548  
 QY 1567 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTGTGCGAGCCCGTGTGACCGCGGTAC 1626  
 Db 1549 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTGTGCGAGCCCGTGTGACCGCGGTAC 1608  
 QY 1627 TAGACCCCGCAGAGGACCTGTGGCGGAGATCCAGAGCAGGGCCACGACCAAGTGGACC 1686  
 Db 1609 TAGACCCCGCAGAGGACCTGTGGCGGAGATCCAGAGCAGGGCCACGACCAAGTGGACC 1668  
 QY 1687 TACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGAGTGGCG 1746  
 Db 1669 TACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGAGTGGCG 1728  
 QY 1747 ACCGCCCCACCAACGACGTGAAGAGCTGACCGAGCCCGTGCAGAAAGATCGCCATGGAG 1806  
 Db 1729 ACCGCCCCACCAACGACGTGAAGAGCTGACCGAGCCCGTGCAGAAAGATCGCCATGGAG 1788  
 QY 1807 AGATGTGTATCTGGGCGACAGACCCCAAGTTCGGCTGCCCCATCCAGAGGAGACTGG 1866  
 Db 1789 AGATGTGTATCTGGGCGACAGACCCCAAGTTCGGCTGCCCCATCCAGAGGAGACTGG 1848  
 QY 1867 GAGACCTGTGGACCGGACTACTTGGCAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1926  
 Db 1849 GAGACCTGTGGACCGGACTACTTGGCAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1908  
 QY 1927 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGAGAGAGAGGCCCATCATCGGCGCCGAG 1986  
 Db 1909 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGAGAGAGAGGCCCATCATCGGCGCCGAG 1968

QY 1987 ACCTTCTAGTGAACGCGCGCCCAACCGCGAGACCAAGATCGCAAGGCGCGCTACGTG 2046  
 Db 1969 ACCTTCTAGTGAACGCGCGCCCAACCGCGAGACCAAGATCGCAAGGCGCGCTACGTG 2028  
 QY 2047 ACCGACCGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAG 2106  
 Db 2029 ACCGACCGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAG 2088  
 QY 2107 CTGACGGCCATCCAGCTGGCCCTGACGAGCAGCGGCGAGGCTGACATCGTGACCGAC 2166  
 Db 2089 CTGACGGCCATCCAGCTGGCCCTGACGAGCAGCGGCGAGGCTGACATCGTGACCGAC 2148  
 QY 2167 AGCAGTGTGCTGACGCGCATCATCCAGGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTG 2226  
 Db 2149 AGCAGTGTGCTGACGCGCATCATCCAGGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTG 2208  
 QY 2227 AACGAGATCATCGAGCAGCTGTATCAAGAGGAGAGTGTACTGAGCTGGTGCCCGCC 2286  
 Db 2209 AACGAGATCATCGAGCAGCTGTATCAAGAGGAGAGTGTACTGAGCTGGTGCCCGCC 2268  
 QY 2287 CACAAGGGCATTCGCGCGCAACGAGCAGATCGCAAGCTGGTGAAGGGCATCCGCAAG 2346  
 Db 2269 CACAAGGGCATTCGCGCGCAACGAGCAGATCGCAAGCTGGTGAAGGGCATCCGCAAG 2328  
 QY 2347 GTGCTGTCTGACGCGCATCGATGCGCGCATCGTGTATCTACAGTACATGGAGCAGCTG 2406  
 Db 2329 GTGCTGTCTGACGCGCATCGATGCGCGCATCGTGTATCTACAGTACATGGAGCAGCTG 2388  
 QY 2407 TAGTGTGGCAGCGCGCCCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 2463  
 Db 2389 TAGTGTGGCAGCGCGCCCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACCGGT 2445

RESULT 12

ADC13264  
 ID ADC13264 standard; DNA; 2445 BP.

XX ADC13264;

XX AC (first entry)

XX 18-DEC-2003 (first entry)

DE DNA of HIV construct p2Pol-opt-YMMW\_C SEQ ID NO 43.  
 XX expression cassette; HIV Gag; Env; Int; Nef; p15NaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX Human immunodeficiency virus.

XX WO2003004620-A2.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

XX 31-AUG-2001; 2001US-0316860P.

XX 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15NaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.

XX Disclosure; Fig 40; 301pp; English.

XX PS

The invention relates to a novel expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p13<sup>gag</sup>, Pol, Tat, Pro, or Rev polypeptide. The novel expression cassette can be used to treat HIV type C by gene therapy or used in the development of a vaccine. The gene delivery vector is administered intramuscularly, intracutaneously, intravenously, subcutaneously, intradermally, transdermally, intravaginally, intrarectally, orally or intravenously. The expression cassette is useful for immunisation, generating packaging cell lines and producing HIV polypeptides. This polynucleotide sequence represents the DNA of an HIV type C related sequence of the invention.

Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 97.3%; Score 2401.8; DB 10; Length 2445;

Best Local Similarity 99.2%; Pred. No. 1.6e-289;

Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

Qy	7	GCACCATGGCGAGGCCATGAGCCAGGCGCACGAGGCCAAATCTGTGATGAGCGCGACG	66
Db	1	GCACCATGGCGAGGCCATGAGCCAGGCGCACGAGGCCAAATCTGTGATGAGCGCGACG	60
Qy	67	AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTTCAACTGCGGCAAGAGAGGGCCACATC	126
Db	61	AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTTCAACTGCGGCAAGAGAGGGCCACATC	120
Qy	127	GCCCGCAATCGCCGCCCCCGCGCAAGAAGGGCTGCTGGAAGTGGCGCAAGAGAGGGCCAC	186
Db	121	GCCCGCAATCGCCGCCCCCGCGCAAGAAGGGCTGCTGGAAGTGGCGCAAGAGAGGGCCAC	180
Qy	187	CAGATGAAGGACTGCACGAGGGCCAGGCGCAACTTTCTTCCGCGAGGACTGCGGCTTCCCG	246
Db	181	CAGATGAAGGACTGCACGAGGGCCAGGCGCAACTTTCTTCCGCGAGGACTGCGGCTTCCCG	240
Qy	247	CAGGGCAAGGCCCGCGAGTTCCCAAGCGAGCAGAAACCGCGCCAAACGCCCCACCAAGCCGC	306
Db	241	CAGGGCAAGGCCCGCGAGTTCCCAAGCGAGCAGAAACCGCGCCAAACGCCCCACCAAGCCGC	300
Qy	307	GAGCTGACAGTGGCGGGCGCAAAACCCCGCAGCGAGGCGCGCGCCGAGCGCCAGGGCACCC	366
Db	301	GAGCTGACAGTGGCGGGCGCAAAACCCCGCGAGCGAGGCGCGCGCCGAGCGCCAGGGCACCC	360
Qy	367	CTGAATCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGGAGCATCAAGGTGGCGCGC	426
Db	361	CTGAATCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGGAGCATCAAGGTGGCGCGC	420
Qy	427	CAGATCAAGGAGGGCCCTGCTGGACACCGCGCGCCGACGACACCGTGTGGAGGAGATGAGC	486
Db	421	CAGATCAAGGAGGGCCCTGCTGGCCACCGCGCGCCGACGACACCGTGTGGAGGAGATGAGC	480
Qy	487	CTGCCCGGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGCTTCTCATCAAGGTGGCG	546
Db	481	CTGCCCGGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGCTTCTCATCAAGGTGGCG	540
Qy	547	CAGTACGACCAAGTCTCTGATCGAGATCTGCGCAAGAAGGCCATCGGACCGCTGCTGATC	606
Db	541	CAGTACGACCAAGTCTCTGATCGAGATCTGCGGCAAGAAGGCCATCGGACCGCTGCTGATC	600
Qy	607	GGCCCCACCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG	666
Db	601	GGCCCCACCCCGTGAAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG	660
Qy	667	AACTTTCCCATCAGGCCCATCGAGACCGTGGCGGTGAAGCTGMAAGCCGGCATGACCGCG	726
Db	661	AACTTTCCCATCAGGCCCATCGAGACCGTGGCGGTGAAGCTGMAAGCCGGCATGACCGCG	720
Qy	727	CCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGGCCCTGACCGCCATCTGC	786
Db	721	CCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGGCCCTGACCGCCATCTGC	780
Qy	787	GAGGATGGAAGAGGGGCAAGATCAACAGATCGGCCCGCGAGAAACCCCTTACAACACC	846
Db	781	GAGGATGGAAGAGGGGCAAGATCAACAGATCGGCCCGCGAGAAACCCCTTACAACACC	840



QY 794 TGGAGAGGGGCAAGATCACCAGATCGGCGCCGAGAACCCCTACCAACACCCCGTGT 853  
 Db 2267 TGGAGAGGGGCAAGATCACCAGATCGGCGCCGAGAACCCCTACCAACACCCCGTGT 2326  
 QY 854 TCGCCATCAAGAAAGAGACAGCAACAAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 913  
 Db 2327 TCGCCATCAAGAAAGAGACAGCAACAAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 2386  
 QY 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGTGGGCATCCCCACCCCGCGGCTGA 973  
 Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGTGGGCATCCCCACCCCGCGGCTGA 2446  
 QY 974 AGAAGAAAGAGCGTGACCGCTGCTGGACGTGGCGAGCGCTTACAGCTGCCCCCTGG 1033  
 Db 2447 AGAAGAAAGAGCGTGACCGCTGCTGGACGTGGCGAGCGCTTACAGCTGCCCCCTGG 2506  
 QY 1034 ACGAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAACAAAGAGACCCCG 1093  
 Db 2507 ACGAGGACTTCCGCAAGTACACCGCTTACCATCCCCAGCATCAACAAAGAGACCCCG 2566  
 QY 1094 GCATCCGCTACAGTACAAAGTGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 1153  
 Db 2567 GCATCCGCTACAGTACAAAGTGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCC 2626  
 QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCT 1213  
 Db 2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCT 2686  
 QY 1214 ACCAGTATGAGCAGCAGCTGTACGTGGGCGAGCGACCTGGAGATGGCGCAGCACCGCGCA 1273  
 Db 2687 ACCA-----GGCCCCCTGTACGTGGGCGAGCGACCTGGAGATGGCGCAGCACCGCGCA 2740  
 QY 1274 AGATCGAGGAGCTGGCGAAGCACCTGTGCTGCGTGGGGCTTCAACACCCCGCAGCAAGAGC 1333  
 Db 2741 AGATCGAGGAGCTGGCGAAGCACCTGTGCTGCGTGGGGCTTCAACACCCCGCAGCAAGAGC 2800  
 QY 1334 ACCAGAGGAGCCCGCTTCTGTGGATGGGTAGCAGCTGCAACCCCGCAGCAAGTGAACG 1393  
 Db 2801 ACCAGAGGAGCCCGCTTCTGTGGATGGGTAGCAGCTGCAACCCCGCAGCAAGTGAACG 2854  
 QY 1394 TGCAGCCATCGAGCTGCGCGAGAGGAGAGCTGAGCGTGAACGATCAGACATCAGAAAGCTGG 1453  
 Db 2855 TGCAGCCATCGAGCTGCGCGAGAGGAGAGCTGAGCGTGAACGATCAGACATCAGAAAGCTGG 2914  
 QY 1454 TGGGCAAGCTGAACTGGGCGAGCCAGATCTACCCCGGCATCAAGTGGCGCCAGCTGTGA 1513  
 Db 2915 TGGGCAAGCTGAACTGGGCGAGCCAGATCTACCCCGGCATCAAGTGGCGCCAGCTGTGA 2974  
 QY 1514 AGCTGCTGCGCGCGCCAAAGCCCTGTACCGACATGTGCTGCCCTGACCGAGAGGCGGAGC 1573  
 Db 2975 AGCTGCTGCGCGCGCCAAAGCCCTGTACCGACATGTGCTGCCCTGACCGAGAGGCGGAGC 3034  
 QY 1574 TGGAGCTGGCGAGAACCGCGAGATCTTGGCGAGCCCGTGCACGGCGTGTACTACGACC 1633  
 Db 3035 TGGAGCTGGCGAGAACCGCGAGATCTTGGCGAGCCCGTGCACGGCGTGTACTACGACC 3094  
 QY 1634 CCAGCAAGGACTGTGTGCGCGAGATCCAGAAAGCAGGGCCAGACAGTGGACCTTACCAGA 1693  
 Db 3095 CCAGCAAGGACTGTGTGCGCGAGATCCAGAAAGCAGGGCCAGACAGTGGACCTTACCAGA 3154  
 QY 1694 TCTACAGGAGCCCTTCAAGAACTGTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCC 1753  
 Db 3155 TCTACAGGAGCCCTTCAAGAACTGTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCC 3214  
 QY 1754 ACACCAAGCAGTGAAGCAGCTGACCGAGGCGGTGCAAGATGCCATGAGAGGACATCG 1813  
 Db 3215 ACACCAAGCAGTGAAGCAGCTGACCGAGGCGGTGCAAGATGCCATGAGAGGACATCG 3274  
 QY 1814 TGATCTGGGCAAGACCCCGCTTCCGCTTCCGCTGCGCATCCAGAAAGAGACCTGGGAGACCT 1873  
 Db 3275 TGATCTGGGCAAGACCCCGCTTCCGCTTCCGCTGCGCATCCAGAAAGAGACCTGGGAGACCT 3334  
 QY 1874 GGTGAGCCGACTACTGGCAGGCCACCTGGATCCCGGAGTGGGAGTTCGTGAACACCCCGCC 1933

Db 3335 GGTGACCGACTACTGGCAGGCGCACTGGATCCCCGAGTGGGAGTTCGTGAACACCCCGCC 3394  
 QY 1934 CCCTGCTGAAGCTGTGTACCAAGCTGGAGAAAGGAGCCCATCATCGCGCGCCGAGACTTCT 1993  
 Db 3395 CCCTGCTGAAGCTGTGTACCAAGCTGGAGAAAGGAGCCCATCATCGCGCGCCGAGACTTCT 3454  
 QY 1994 ACGTGGACGGCGCGCCCAACCCGCGAGACCAAGATCGGCAAGGCCCGGCTACGTGACCGACC 2053  
 Db 3455 ACGTGGACGGCGCGCCCAACCCGCGAGACCAAGATCGGCAAGGCCCGGCTACGTGACCGACC 3514  
 QY 2054 GGGGCGGCGCAAGATCGTGAAGCCCTGACCGAGACCCACCAACAGAAAGCCGAGCTGCAGG 2113  
 Db 3515 GGGGCGGCGCAAGATCGTGAAGCCCTGACCGAGACCCACCAACAGAAAGCCGAGCTGCAGG 3574  
 QY 2114 CCATCCAGCTGGCCCTGCGAGCAGCGAGCGAGGTGAACATCGTGACCCGACAGCCAGT 2173  
 Db 3575 CCATCCAGCTGGCCCTGCGAGCAGCGAGCGAGGTGAACATCGTGACCCGACAGCCAGT 3634  
 QY 2174 ACGCCCTGGGCATCATTCAGGCCCAAGCCCGCAGAGCGAGCGAGCTGTGTGAACAGAG 2233  
 Db 3635 ACGCCCTGGGCATCATTCAGGCCCAAGCCCGCAGAGCGAGCGAGCTGTGTGAACAGAG 3694  
 QY 2234 TCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACCTGAGCTGGGTGCGCCCGCCCAAGG 2293  
 Db 3695 TCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACCTGAGCTGGGTGCGCCCGCCCAAGG 3754  
 QY 2294 GCATCGGCGGCAAGCAGCAGATCGAACAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353  
 Db 3755 GCATCGGCGGCAAGCAGCAGATCGAACAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 3814  
 QY 2354 TCCTGGACGGCATCGATGGCGGCATCGGTGATCTACAGTACATGGAACGACCTGTACGTGG 2413  
 Db 3815 TCCTGGACGGCATCGATGGCGGCATCGGTGATCTACAGTACATGGAACGACCTGTACGTGG 3874  
 QY 2414 GCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCCGT 2463  
 Db 3875 GCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCCGT 3924  
 RESULT 14  
 ADCL13231  
 ID ADCL13231 standard; DNA; 3930 BP.  
 XX  
 AC ADCL13231;  
 XX AC  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE DNA of HIV construct GagComplPolmutAtt\_C SEQ ID NO 10.  
 XX  
 KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 FN WO2003004620-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021420.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (UYST-) UNIV STELLENBOSCH.  
 XX  
 PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 XX  
 DR WPI; 2003-221593/21.  
 XX  
 PT New expression cassette comprising a polynucleotide sequence encoding a

PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
PS Disclosure; Fig 7; 301pp; English.  
XX  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscularly, intravenously,  
CC subcutaneously, intradermally, intradermally, intravenously,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC type C related sequence of the invention.  
XX  
SQ Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;  
Query Match 96.98; Score 2393.2; DB 10; Length 3930;  
Best Local Similarity 99.28; Pred. No. 1.7e-288;  
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;  
QY 14 TGGCCGAGGCGCATGAGCCAGCCACCAAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA 73  
DB 1487 TCGCCGAGGCGCATGAGCCAGCCACCAAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA 1546  
QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTTGGCGCAAGGAGGGCCACATCGCCCGCA 133  
DB 1547 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTTGGCGCAAGGAGGGCCACATCGCCCGCA 1606  
QY 134 ACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGAGGGCGCCACAGATGA 193  
DB 1607 ACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGCAGGAGGGCGCCACAGATGA 1666  
QY 194 AGGACTGACACGAGCGCCAGCCCAACTTCTTCCGCGAGGACCTTGGCCCTTCCCCCAGGGCA 253  
DB 1667 AGGACTGACACGAGCGCCAGCCCAACTTCTTCCGCGAGGACCTTGGCCCTTCCCCCAGGGCA 1726  
QY 254 AGGCGCGCGAGTTCCCGAGCGAGAGAACCCGCGCAACAGCCCGCAACAGCCCGGAGCTGC 313  
DB 1727 AGGCGCGCGAGTTCCCGAGCGAGAGAACCCGCGCAACAGCCCGGAGCTGC 1786  
QY 314 AGGTGCGCGCGCAACCCCGCAGCGAGGCGCGCGCGCGAGCGCGCAGGGCACCTGAACCT 373  
DB 1787 AGGTGCGCGCGCAACCCCGCAGCGAGGCGCGCGCGCGAGCGCGCAGGGCACCTGAACCT 1846  
QY 374 TCCCCCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGGTGGCGCGCCAGATCA 433  
DB 1847 TCCCCCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGGTGGCGCGCCAGATCA 1906  
QY 434 AGGAGGCCCTGTGGACACCGCGCGCGCAACACACCGTGTGTGAGGAGATGAGCTGCGCCG 493  
DB 1907 AGGAGGCCCTGTGGACTCTCGCGCGCGCAACACACCGTGTGTGAGGAGATGAGCTGCGCCG 1966  
QY 494 GCAAGTGGAGGCCCAAGATCATCGCGCGCATCGCGCGCTTTCATCAAGGTGCGCGCAGTACG 553  
DB 1967 GCAAGTGGAGGCCCAAGATCATCGCGCGCATCGCGCGCTTTCATCAAGGTGCGCGCAGTACG 2026  
QY 554 ACCAGATCTCTGATCGAGATCTGGGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA 613  
DB 2027 ACCAGATCTCTGATCGAGATCTGGGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA 2086  
QY 614 CCCCCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGTGACCCCTGAATCTTCC 673  
DB 2087 CCCCCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGTGACCCCTGAATCTTCC 2146  
QY 674 CCATCAGCCCCCATCGAGACCGTGCCTGTGAAGTGTGAAGCCCGCGCATGAGCGGCCCAAGG 733  
DB 2147 CCATCAGCCCCCATCGAGACCGTGCCTGTGAAGTGTGAAGCCCGCGCATGAGCGGCCCAAGG 2206  
QY 734 TGAAGCAGTGGCCCCCTGACCCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGCGAGGAGA 793

DB 2207 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGA 2266  
QY 794 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTACAACACACCCCGTGT 853  
DB 2267 TGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTACAACACCCCGTGT 2326  
QY 854 TCGCCCATCAAGAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGTGTGGACTTCCCGGAGCTGA 913  
DB 2327 TCGCCCATCAAGAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGTGTGGACTTCCCGGAGCTGA 2386  
QY 914 ACAAGCGCACCCAGGACTTCTTGGAGGATGAGCTGGGCAATCCCCACACCCCGCGGCTTGA 973  
DB 2387 ACAAGCGCACCCAGGACTTCTTGGAGGATGAGCTGGGCAATCCCCACACCCCGCGGCTTGA 2446  
QY 974 AGAGAGAGAGAGCGGTGACCGTGTGAGCGTGGCGAGCGCTACTTTCAGCGTGCCTTGG 1033  
DB 2447 AGAGAGAGAGAGCGGTGACCGTGTGAGCGTGGCGAGCGCTACTTTCAGCGTGCCTTGG 2506  
QY 1034 ACAGGACTTTCGCGAAAGTACACCGCCTTCAACATCCCCAGCATCAACAAACGAGACCCCG 1093  
DB 2507 ACAGGACTTTCGCGAAAGTACACCGCCTTCAACATCCCCAGCATCAACAAACGAGACCCCG 2566  
QY 1094 GCATCCGCTACCAAGTACAAGTGTCTGCCCGAGGGCTGGAAGGGCAGGCCCGAGCATCTTCC 1153  
DB 2567 GCATCCGCTACCAAGTACAAGTGTCTGCCCGAGGGCTGGAAGGGCAGGCCCGAGCATCTTCC 2626  
QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCT 1213  
DB 2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCT 2686  
QY 1214 ACCAGTACATGAGCAGACCTGTAGCTGGGAGCAGCACCTGGAGATCGGCGAGCACCGGCCA 1273  
DB 2687 ACCA-----GGCCCCCTGTAGCTGGGAGCAGCACCTGGAGATCGGCGAGCACCGGCCA 2740  
QY 1274 AGATCAGAGAGTGCAGCAAGCAGCTGTGTGTGGGCTTCAACACCCCGCAACAAGAGC 1333  
DB 2741 AGATCAGAGAGTGCAGCAAGCAGCTGTGTGTGGGCTTCAACACCCCGCAACAAGAGC 2800  
QY 1334 ACCAGAGAGAGCCCCCTTCTGTGTGGGCTTCAACACCCCGCAACAAGAGTGGACCG 1393  
DB 2801 ACCAGAGAGAGCCCCCTTCTGTGTGGGCTTCAACACCCCGCAACAAGTGGACCG 2854  
QY 1394 TGCAGGCCCATCGAGCTGCCCGAGAGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGG 1453  
DB 2855 TGCAGGCCCATCGAGCTGCCCGAGAGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGG 2914  
QY 1454 TGGGCAAGCTGAACCTGGGCGAGCAGATCTAACCAGGATCAAGGTGCGCGAGCTGTGCA 1513  
DB 2915 TGGGCAAGCTGAACCTGGGCGAGCAGATCTAACCAGGATCAAGGTGCGCGAGCTGTGCA 2974  
QY 1514 AGCTGTGCGCGCGCCAGAGCCCTGACACATCGTGGCCCTGACCGAGGAGGCCGAGC 1573  
DB 2975 AGCTGTGCGCGCGCCAGAGCCCTGACCGACATCGTGGCCCTGACCGAGGAGGCCGAGC 3034  
QY 1574 TGCAGCTGGCGCGAGAACCCGAGATCTTGGCGAGCCGCTGACCGGCTGTACTACGACC 1633  
DB 3035 TGCAGCTGGCGCGAGAACCCGAGATCTTGGCGAGCCGCTGACCGGCTGTACTACGACC 3094  
QY 1634 CAGCAGAGCAGCTGTGGTGGCGAGATCCAGAGAGAGGCGCACGACGATGAGACCTTACAGA 1693  
DB 3095 CCAGCAGAGCAGCTGTGGTGGCGAGATCCAGAGAGAGGCGCACGACGATGAGACCTTACAGA 3154  
QY 1694 TCTTACAGGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCAAGATCGCGACCGCC 1753  
DB 3155 TCTTACAGGAGCCCTTCAAGAACCTGAGAACCGGCAAGTACGCAAGATCGCGACCGCC 3214  
QY 1754 ACACCAACGAGCTGAAGCAGCTGACCGAGCCGCTGAGAGAGATCGCCATGAGAGCATCG 1813  
DB 3215 ACACCAACGAGCTGAAGCAGCTGACCGAGCCGCTGAGAGAGATCGCCATGAGAGCATCG 3274  
QY 1814 TGAATCTGGGCGCAAGACCCCAAGTTTCGCTTGGCCATCCAGAGGAGACCTTGGAGACCT 1873

Db 3275 TGATCTGGGCAAGACCCCAAGTTCGCTGCCATCCAGAAGAGACCTGGAGACCT 3334  
Qy 1874 GGTGACCGACTACTGGCAGCCCACTGGATCCCGAGTGGAGTTGCTGAACACCCCCC 1933  
Db 3335 GGTGACCGACTACTGGCAGCCCACTGGATCCCGAGTGGAGTTGCTGAACACCCCCC 3394  
Qy 1934 CCTGTGTAGCTGTGTACCACTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCT 1993  
Db 3395 CCTGTGTAGCTGTGTACCACTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCT 3454  
Qy 1994 ACCTGTGACCGCGCCGCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACCGACC 2053  
Db 3455 ACCTGTGACCGCGCCGCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTGACCGACC 3514  
Qy 2054 GGGCGCGGCAAGAGATCGTGTAGCTGTGACCGAGACCAACCAAGAGACCGAGTGCAGG 2113  
Db 3515 GGGCGCGGCAAGAGATCGTGTAGCTGTGACCGAGACCAACCAAGAGACCGAGTGCAGG 3574  
Qy 2114 CCATCCAGCTGGCCCTGCAGACAGCGGACGAGGTGMACTGTGACCGACAGCCAGT 2173  
Db 3575 CCATCCAGCTGGCCCTGCAGACAGCGGACGAGGTGMACTGTGACCGACAGCCAGT 3634  
Qy 2174 ACGCCCTGGGCATCATCCAGGCCCGAGCCGCAAGAGCGAGCGAGTGGTGAACAGA 2233  
Db 3635 ACGCCCTGGGCATCATCCAGGCCCGAGCCGCAAGAGCGAGCGAGTGGTGAACAGA 3694  
Qy 2234 TCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTTGTGAGTGGGTGCCGCCCAAGG 2293  
Db 3695 TCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTTGTGAGTGGGTGCCGCCCAAGG 3754  
Qy 2294 GCATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGCATCCGCAAGTGTGT 2353  
Db 3755 GCATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGCATCCGCAAGTGTGT 3814  
Qy 2354 TCCTGAGCGCATCGATGGCGGCATCGTGTACTACCACTGATGAGCAGCTGTACCTGG 2413  
Db 3815 TCCTGAGCGCATCGATGGCGGCATCGTGTACTACCACTGATGAGCAGCTGTACCTGG 3874  
Qy 2414 GCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463  
Db 3875 GCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

## RESULT 15

ADC13232  
ID ADC13232 standard; DNA; 3930 BP.

XX AC ADC13232;

XX DX 18-DEC-2003 (first entry)

XX DE DNA of HIV construct GagCompIPolmutina\_c SEQ ID NO 11.

XX DX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
XX KW Rev; HIV type C; Gene therapy; vaccine; immunisation; HIV; ds.

XX OS Human immunodeficiency virus.

XX PN W02003004620-A2.

XX PD 16-JAN-2003.

XX PF 05-JUL-2002; 2002WO-US021420.

XX PR 05-JUL-2001; 2001US-0303192P.

XX PR 31-AUG-2001; 2001US-0316860P.

XX PR 16-JAN-2002; 2002US-0349871P.

XX PA (CHIR ) CHIRON CORP.

XX PA (UYST-) UNIV STELLENBOSCH.

XX PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX XX

DR WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.

XX Disclosure; Fig 8; 301pp; English.

PS The invention relates to a novel expression cassette comprising a  
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramuscularly, intravenously,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.

XX SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 96.9%; Score 2393.2; DB 10; Length 3930;

Best Local Similarity 99.2%; Pred. No. 1.7e-288;

Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

Qy 14 TGGCCGAGGCGCATGAGCGAGCCACAGCGCCCAACATCTCTGATCGAGCGCAACTTCA 73

Db 1487 TCGCCGAGGCGCATGAGCGAGCCACAGCGCCCAACATCTCTGATCGAGCGCAACTTCA 1546

Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTTCAACTGCGGCAAGAGGGCCACATCCCGCGCA 133

Db 1547 AGGCCCCCAAGCGCATCATCAAGTGTTCAACTGCGGCAAGAGGGCCACATCCCGCGCA 1606

Qy 134 ACTGCCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAGAGGGCCACAGATGA 193

Db 1607 ACTGCCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAGAGGGCCACAGATGA 1666

Qy 194 AGGACTGCACCGCGCGCAGCGCCCAACTTCTTCCCGGAGGACCTGGCCCTTCCCCCAGGGCA 253

Db 1667 AGGACTGCACCGCGCGCAGCGCCCAACTTCTTCCCGGAGGACCTGGCCCTTCCCCCAGGGCA 1726

Qy 254 AGGCCCCGAGTTCCCGCAGCGAGCAGAAACCGCGCAACAGCCCCCAGCGCGCGAGTGC 313

Db 1727 AGGCCCCGAGTTCCCGCAGCGAGCAGAAACCGCGCAACAGCCCCCAGCGCGAGTGC 1786

Qy 314 AGGTGCGCGGAGACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373

Db 1787 AGGTGCGCGGAGACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846

Qy 374 TCCCGCCAGATCACCTGTGCGAGCGCCCTCTGTGTGAGCATCAAGTGGCGCGCGCGCGCGCA 433

Db 1847 TCCCGCCAGATCACCTGTGCGAGCGCCCTCTGTGTGAGCATCAAGTGGCGCGCGCGCGCA 1906

Qy 434 AGGAGGCCCTGTCTGGACACCG 493

Db 1907 AGGAGGCCCTGTCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966

Qy 494 GCAAGTGAAGCCCAAGATGATCGGCGCGCATCGGCGCGCTTCAATCAAGTGGCGCGCGCGCG 553

Db 1967 GCAAGTGAAGCCCAAGATGATCGGCGCGCATCGGCGCGCTTCAATCAAGTGGCGCGCGCG 2026

Qy 554 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCGCTTCAATCGGCGCGCA 613

Db 2027 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCGCTTCAATCGGCGCGCA 2086

Qy 614 CCCCCTGGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGGAATTCC 673

Db 2087 CCCCCTGGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGGAATTCC 2146

Qy 674 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGCGCATGAGCGCGCGCGCGCG 733

XX



Db 2147 CCATCAGCCCCATCGAGACGTCGTCGCGTGAAGCTGAAGCCCGGCATGAGACGGCCCCAAGG 2206  
Qy 734 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGCCCTGACCGGCATCTGCGAGAGA 793  
Db 2207 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGCCCTGACCGGCATCTGCGAGAGA 2266  
Qy 794 TGAAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAACCCCTTCAACACCCCGCTGT 853  
Db 2267 TGAAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAACCCCTTCAACACCCCGCTGT 2326  
Qy 854 TGCCTCATCAAGAAAGAGAGACAGACCAAGTGGCGCAAGCTGTGTGAGATCTTCGGGAGCTGA 913  
Db 2327 TGCCTCATCAAGAAAGAGAGACAGACCAAGTGGCGCAAGCTGTGTGAGATCTTCGGGAGCTGA 2386  
Qy 914 ACAAGCGCACCCAGGATCTCTGGAGGTGTCAGCTGGGCATCCCCACCGCCGGCCTGA 973  
Db 2387 ACAAGCGCACCCAGGATCTCTGGAGGTGTCAGCTGGGCATCCCCACCGCCGGCCTGA 2446  
Qy 974 AGAAGAGAGAGCGTGTGACCGTGTCTGGACGTGGGCGACGCTACTTTCAGCGTGCCTCTGG 1033  
Db 2447 AGAAGAGAGAGCGTGTGACCGTGTCTGGACGTGGGCGACGCTACTTTCAGCGTGCCTCTGG 2506  
Qy 1034 ACAGAGCTTCGCAAGATGACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCG 1093  
Db 2507 ACAGAGCTTCGCAAGATGACCGCTTCAACATCCCGAGCATCAACAAACGAGACCCCG 2566  
Qy 1094 GCATCCGCTACAGTACAAAGTGTGCTGCCAGGCTGGAAGGCGACGCCAGCATCTTC 1153  
Db 2567 GCATCCGCTACAGTACAAAGTGTGCTGCCAGGCTGGAAGGCGACGCCAGCATCTTC 2626  
Qy 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213  
Db 2627 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2686  
Qy 1214 ACCAGTACATGAGACGACTGTGTCGCGGACGCACTGTGAGATCGGCGAGCACCGGCCA 1273  
Db 2687 ACCA-----GGCCCCCTGTACGTGGGACGACCTGTGAGATCGGCGACACCGGCCA 2740  
Qy 1274 AGATCAGGAGCTGCCAGACACTCTGCGCTGGGGCTTCAACCCCGCAAGAAGC 1333  
Db 2741 AGATCAGGAGCTGGCGAAGCACTCTGCGCTGGGGCTTCAACCCCGCAAGAAGC 2800  
Qy 1334 ACCAGAAGAGACCCCTCTCTGTGATGGGCTACGAGCTGCAACCCCGCAAGTGGACCG 1393  
Db 2801 ACCAGAAGAGACCCCTCTCTGTGATGGGCTACGAGCTGCAACCCCGCAAGTGGACCG 2854  
Qy 1394 TGCAGCCCATCGAGCTGCCGAGAGAGAGCTGAGCCGTGGAACGACATCCAGAACCTGG 1453  
Db 2855 TGCAGCCCATCGAGCTGCCGAGAGAGAGCTGAGCCGTGGAACGACATCCAGAACCTGG 2914  
Qy 1454 TGGGCAAGCTGAATGGGCGCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 1513  
Db 2915 TGGGCAAGCTGAATGGGCGCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCA 2974  
Qy 1514 AGCTGTGCGCGCGCCCAAGCCCTGACCGACATCGTGCCTCTGACCGAGAGCGCGAGC 1573  
Db 2975 AGCTGTGCGCGCGCCCAAGCCCTGACCGACATCGTGCCTCTGACCGAGAGCGCGAGC 3034  
Qy 1574 TGAAGTGGCCGAGAACCGCGAGATCTGTGCGAGGCCGTGTCACGGCGTGTACTAGACC 1633  
Db 3035 TGAAGTGGCCGAGAACCGCGAGATCTGTGCGAGGCCGTGTCACGGCGTGTACTAGACC 3094  
Qy 1634 CCAGCAGGACCTGGTGGCCGAGATCCAGAGCAGGGCCACGACCTGAGTGGACCTTACCAGA 1693  
Db 3095 CCAGCAGGACCTGGTGGCCGAGATCCAGAGCAGGGCCACGACCTGAGTGGACCTTACCAGA 3154  
Qy 1694 TCTACAGGAGCCCTTCAAGAACTGAAAGCCGCGCAAGTACGCGCAAGATCGCAACCGCCC 1753  
Db 3155 TCTACAGGAGCCCTTCAAGAACTGAAAGCCGCGCAAGTACGCGCAAGATCGCAACCGCCC 3214  
Qy 1754 ACACCAACGAGCTGAGAGCTGACCGAGCCGTGCGAAGATCGCCATGAGAGCATCG 1813  
Db 3215 ACACCAACGAGCTGAGAGCTGACCGAGCCGTGCGAAGATCGCCATGAGAGCATCG 3274

Qy 1814 TGATCTGGGGCAAGACCCCAAGTTTCCGCTGCTCCATCCAGAAAGAGACCTGGGAGACCT 1873  
Db 3275 TGATCTGGGGCAAGACCCCAAGTTTCCGCTGCTCCATCCAGAAAGAGACCTGGGAGACCT 3334  
Qy 1874 GGTGAGACCACTACTGCGCAGGCGACCTGATCCCGAGTGGGAGTTTCTGTGAACACCCCC 1933  
Db 3335 GGTGAGACCACTACTGCGCAGGCGACCTGATCCCGAGTGGGAGTTTCTGTGAACACCCCC 3394  
Qy 1934 CCTGTGTAAGCTGTGTTACAGCTGAGGAAGAGCCCATCATCGGCGCCGAGACCTTCT 1993  
Db 3395 CCTGTGTAAGCTGTGTTACAGCTGAGGAAGAGGCCCATCATCGGCGCCGAGACCTTCT 3454  
Qy 1994 AGTGTGACCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 2053  
Db 3455 AGTGTGACCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACC 3514  
Qy 2054 GGGCGCGCAGAAAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGG 2113  
Db 3515 GGGCGCGCAGAAAGATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGG 3574  
Qy 2114 CCATCCAGCTGGCCCTGACAGGACAGCGGACGAGGTGAACATCTGTGACCGAGCCAGT 2173  
Db 3575 CCATCCAGCTGGCCCTGACAGGACAGCGGACGAGGTGAACATCTGTGACCGAGCCAGT 3634  
Qy 2174 AGCCCTGGGCGCATCATCCAGGCCCCAGCCGACAAAGAGCGAGCGAGCTGGTGAAACAGA 2233  
Db 3635 AGCCCTGGGCGCATCATCCAGGCCCCAGCCGACAAAGAGCGAGCGAGCTGGTGAAACAGA 3694  
Qy 2234 TCATCAGCAGCTGATCAAGAAAGAGAGAGTGTACTGTAGCTGGGTGCCGCCACAGG 2293  
Db 3695 TCATCAGCAGCTGATCAAGAAAGAGAGAGTGTACTGTAGCTGGGTGCCGCCACAGG 3754  
Qy 2294 GCATCGCGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCGATCCGCAAGGTGCTGT 2353  
Db 3755 GCATCGCGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCGATCCGCAAGGTGCTGT 3814  
Qy 2354 TCCTGACCGGCATTCGATGGCGGCATCTGTATCTTACCAGTACATGAGACGACCTGTACGTGG 2413  
Db 3815 TCCTGACCGGCATTCGATGGCGGCATCTGTATCTTACCAGTACATGAGACGACCTGTACGTGG 3874  
Qy 2414 GCMGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463  
Db 3875 GCAGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

Search completed: June 1, 2005, 11:33:18  
Job time : 1280.42 secs





Db 181 GGCCACAGATGAAGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240  
Qy 241 TTCCCCAGGCGAAGGCCCGCGAGTTCCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACC 300  
Db 241 TTCCCCAGGCGAAGGCCCGCGAGTTCCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACC 300  
Qy 301 AGCGCGAGCTGCGAGGTGCGCGCGCAACACCCCGCAGCAGGCGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTGCGAGGTGCGCGCGCAACACCCCGCAGCAGGCGCGCGCGCGAGCGCCAG 360  
Qy 361 GGCACTGTAACTTCCCCAGATCACTGTGTGAGCGCCCCCTGTGTAGCATCAAGGTG 420  
Db 361 GGCACTGTAACTTCCCCAGATCACTGTGTGAGCGCCCCCTGTGTAGCATCAAGGTG 420  
Qy 421 GCGGCGCAGATCAAGAGGCCCTGTGTGAACCGGCGCCGACACACCTGTGTGGAGGAG 480  
Db 421 GCGGCGCAGATCAAGAGGCCCTGTGTGAACCGGCGCCGACACACCTGTGTGGAGGAG 480  
Qy 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGCGCTTTCATCAAG 540  
Db 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGCGCTTTCATCAAG 540  
Qy 541 GTGCGGCAGTACACCAAGATCTCTGATCGAGATCTGTGGGCAAGAGCCATCGGCACCGTG 600  
Db 541 GTGCGGCAGTACACCAAGATCTCTGATCGAGATCTGTGGGCAAGAGCCATCGGCACCGTG 600  
Qy 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTTGGGCTGC 660  
Db 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTTGGGCTGC 660  
Qy 661 ACCCTGAATTTCCCATCAGCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
Db 661 ACCCTGAATTTCCCATCAGCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
Qy 721 GACGGCCCCAAGGTGAAGTAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCC 780  
Db 721 GACGGCCCCAAGGTGAAGTAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCC 780  
Qy 781 ATCTGGAGAGATGGAAGAGAGGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTAC 840  
Db 781 ATCTGGAGAGATGGAAGAGAGGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTAC 840  
Qy 841 AACACCCCGCTGTTCGCCATCAAGAGAGGACAGACCAAGTGGCGCAAGCTGTGTGAC 900  
Db 841 AACACCCCGCTGTTCGCCATCAAGAGAGGACAGACCAAGTGGCGCAAGCTGTGTGAC 900  
Qy 901 TTCGGGAGCTGAACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCATCCCCCAC 960  
Db 901 TTCGGGAGCTGAACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCATCCCCCAC 960  
Qy 961 CCGCGGCTGAAGAGAGAGGGTGAACCGTGTGACCTGGCGAGCGCTACTTC 1020  
Db 961 CCGCGGCTGAAGAGAGAGGGTGAACCGTGTGACCTGGCGAGCGCTACTTC 1020  
Qy 1021 AGCGTCCCTGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
Db 1021 AGCGTCCCTGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCCCCAGGGCTGGAAGGGGAGC 1140  
Db 1081 AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCCCCAGGGCTGGAAGGGGAGC 1140  
Qy 1141 CCCAGCATTTCCAGAGCAGATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200  
Db 1141 CCCAGCATTTCCAGAGCAGATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200  
Qy 1201 GAGATCGTGTATCAACAGTACATGGACACCTGTGCTGGGCGAGCGACCTGGAGATCGGC 1260  
Db 1201 GAGATCGTGTATCAACAGTACATGGACACCTGTGCTGGGCGAGCGACCTGGAGATCGGC 1260  
Qy 1261 CAGCACCGGCCAAGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCAACACC 1320  
Db 1261 CAGCACCGGCCAAGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCAACACC 1320

Qy 1321 CCGCACAAGAAGCACACAGAAGGAGCCCCCTTCTGTGGATGGCTACGAGCTGCACCCC 1380  
Db 1321 CCGCACAAGAAGCACACAGAAGGAGCCCCCTTCTGTGGATGGCTACGAGCTGCACCCC 1380  
Qy 1381 GACAAGTGGACCCGTGACGCCATCGAGCTGCCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440  
Db 1381 GACAAGTGGACCCGTGACGCCATCGAGCTGCCCCGAGAAGGAGAGCTGGACCGTGAACGAC 1440  
Qy 1441 ATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500  
Db 1441 ATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG 1500  
Qy 1501 CGCCAGCTGTGCAAGCTGTGCGCGGCCAGCGCTGACGACATCTGTGCCCTTGACC 1560  
Db 1501 CGCCAGCTGTGCAAGCTGTGCGCGGCCAGCGCTGACGACATCTGTGCCCTTGACC 1560  
Qy 1561 GAGGAGGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTGTGCGGAGCCCTGTGCACGCG 1620  
Db 1561 GAGGAGGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTGTGCGGAGCCCTGTGCACGCG 1620  
Qy 1621 GTGTACTACGACCCCGAGAAAGGACCTGTGTGGCCGAGATCCAGAAGCAGGGGCCAGCCAG 1680  
Db 1621 GTGTACTACGACCCCGAGAAAGGACCTGTGTGGCCGAGATCCAGAAGCAGGGGCCAGCCAG 1680  
Qy 1681 TGGACCTTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCCGCAAGTACGCCAAG 1740  
Db 1681 TGGACCTTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCCGCAAGTACGCCAAG 1740  
Qy 1741 ATGCGCACCGCCCAACCAACGACGTGAAGACGTGACCGAGGCCGTGCAAGAGATCGCC 1800  
Db 1741 ATGCGCACCGCCCAACCAACGACGTGAAGACGTGACCGAGGCCGTGCAAGAGATCGCC 1800  
Qy 1801 ATGGAAGCATCTGTATCTGGGCAAGACCCCAAGTTCCGCCCTGCCATCCAGAAGGAG 1860  
Db 1801 ATGGAAGCATCTGTATCTGGGCAAGACCCCAAGTTCCGCCCTGCCATCCAGAAGGAG 1860  
Qy 1861 ACCTGGAGACCTGTGTGGACCGGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTC 1920  
Db 1861 ACCTGGAGACCTGTGTGGACCGGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTC 1920  
Qy 1921 GTGAACAACCCCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAAGGAGCCCATCATCGC 1980  
Db 1921 GTGAACAACCCCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAAGGAGCCCATCATCGC 1980  
Qy 1981 GCGGAGACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCCGCAAGGCCGCG 2040  
Db 1981 GCGGAGACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCCGCAAGGCCGCG 2040  
Qy 2041 TAGCTGACCGACCGGGCGCGGCAAGATCGTGAAGCTGTGTGTACAGCTGACCGAGACCAACCAAG 2100  
Db 2041 TAGCTGACCGACCGGGCGCGGCAAGATCGTGAAGCTGTGTGTACCGAGACCAACCAAG 2100  
Qy 2101 ACCGAGCTGACGCCATCCAGCTGGCCCTGACGAGACGCGGAGCGAGGTGAACATCGTG 2160  
Db 2101 ACCGAGCTGACGCCATCCAGCTGGCCCTGACGAGACGCGGAGCGAGGTGAACATCGTG 2160  
Qy 2161 ACCGACAGCCAGTACGCGCTGGGCGATCATCCAGGCCCGCCGACAGAGGAGGAGCGAG 2220  
Db 2161 ACCGACAGCCAGTACGCGCTGGGCGATCATCCAGGCCCGCCGACAGAGGAGGAGCGAG 2220  
Qy 2221 CTGGTGAAACAGATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2280  
Db 2221 CTGGTGAAACAGATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2280  
Qy 2281 CCGCCCCACAAGGCGATCGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATC 2340  
Db 2281 CCGCCCCACAAGGCGATCGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATC 2340  
Qy 2341 CGCAAGGTGCTGTCTCTGAGCGGATCGATGGGGCATCGTGTATCTACCAAGTACATGGAC 2400  
Db 2341 CGCAAGGTGCTGTCTCTGAGCGGATCGATGGGGCATCGTGTATCTACCAAGTACATGGAC 2400

Qy 2401 GACCTGACGTGGGAGGCGGCGCTAGGATCGATTAAAGCTTCCGGGGCTAGCAC 2460  
Db 2401 GACCTGACGTGGGAGGCGGCGCTAGGATCGATTAAAGCTTCCGGGGCTAGCAC 2460  
Qy 2461 GGTGAATTC 2469  
Db 2461 GGTGAATTC 2469

## RESULT 2

AX455915 2463 bp DNA linear PAT 06-JUL-2002  
LOCUS  
DEFINITION Sequence 31 from Patent W00204493.  
ACCESSION AX455915  
VERSION AX455915.1 GI:21714900  
KEYWORDS  
SOURCE  
ORGANISM

synthetic construct  
synthetic construct  
other sequences; artificial sequences.

## REFERENCE

1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.

Poly nucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
Patent: WO 0204493-A 31 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)

## JOURNAL

Location/Qualifiers

## FEATURES

source  
1. 2463  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YM"

## ORIGIN

Query Match 98.9%; Score 2442.2; DB 6; Length 2463;

Best Local Similarity 99.6%; Pred. No. 1.6e-246;

Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTCAGCCCACTATGGCCGAGGCGCATGAGCAGGCCACAGCGCCAACTCTGATGAG 60  
Db 1 GTCAGCCCACTATGGCCGAGGCGCATGAGCAGGCCACAGCGCCAACTCTGATGAG 60  
Qy 61 CGCAGCACTTCAGGGCCCGCCAGCGCATCATCAAGTGTCTCACTGGCGGAGGGC 120  
Db 61 CGCAGCACTTCAGGGCCCGCCAGCGCATCATCAAGTGTCTCACTGGCGGAGGGC 120  
Qy 121 CACATCCCGCCCACTGCGCGGCGCCCGCAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCCCGCCCACTGCGCGGCGCCCGCAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Qy 181 GGCACCCAGATGAAGGACTGACCGAGCGCGCCAGCCAACTTCTTCGCGAGGACCTGGCC 240  
Db 181 GGCACCCAGATGAAGGACTGACCGAGCGCGCCAGCCAACTTCTTCGCGAGGACCTGGCC 240  
Qy 241 TTTCCCGCAGGCGCGCGGCGCAACCCCGCAGCGAGGCGCGCGCCCAAGCCCGCACC 300  
Db 241 TTTCCCGCAGGCGCGCGGCGCAACCCCGCAGCGAGGCGCGCGCCCAAGCCCGCACC 300  
Qy 301 AGCGCGAGCTGCAAGTGTGCGGGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGCAG 360  
Db 301 AGCGCGAGCTGCAAGTGTGCGGGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGCAG 360  
Qy 361 GGCAACCTTGAATTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGTG 420  
Db 361 GGCAACCTTGAATTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGTG 420  
Qy 421 GCGCGCAGATCAAGGAGGCGCTGCTGACACCGCGCGCGCGCGCGCGCGCGCGGAG 480  
Db 421 GCGCGCAGATCAAGGAGGCGCTGCTGACACCGCGCGCGCGCGCGCGCGCGGAG 480  
Qy 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGGCTTTCATCAAG 540  
Db 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGGCTTTCATCAAG 540

Qy 541 GTGCGCCAGTACGACGAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGACCGGTG 600  
Db 541 GTGCGCCAGTACGACGAGATCTGATCGAGATCTGCGGCAAGAGGCCATCGGACCGGTG 600  
Qy 601 CTGATCGGCGCCACCCCGCTGAAACATCATCGGCGGCAACATGCTGACAGCTGGGCTGC 660  
Db 601 CTGATCGGCGCCACCCCGCTGAAACATCATCGGCGGCAACATGCTGACAGCTGGGCTGC 660  
Qy 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGATG 720  
Qy 721 GACGCGCCCAAGTGAAGCAGTGGCCCTGACCGAGGAGATCAAGGCGCTGACCGCC 780  
Db 721 GACGCGCCCAAGTGAAGCAGTGGCCCTGACCGAGGAGATCAAGGCGCTGACCGCC 780  
Qy 781 ATCTGCGAGGAGTGAAGAGGAGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGTGAAGAGGAGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTAC 840  
Qy 841 AACACCCCGCTGTTCCCATCAAGAGAGGAGCAGCACCAGTGGCGCAAGCTGGTGAC 900  
Db 841 AACACCCCGCTGTTCCCATCAAGAGAGGAGCAGCACCAGTGGCGCAAGCTGGTGAC 900  
Qy 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGCC 960  
Db 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGCC 960  
Qy 961 CCGCGCGGCTGAAAGAGAGAGAGGCGTGAACCGTGTGAGCGTGGGCGAGCGCTACTTC 1020  
Db 961 CCGCGCGGCTGAAAGAGAGAGAGGCGTGAACCGTGTGAGCGTGGGCGAGCGCTACTTC 1020  
Qy 1021 AGCGTGCCTTGGACGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGCCTTGGACGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGCTACCGTACAGTACAACTGCTGCCCGCAGGCTGGAAGGGGAGC 1140  
Db 1081 AACGAGACCCCGCGCTACCGTACAGTACAACTGCTGCCCGCAGGCTGGAAGGGGAGC 1140  
Qy 1141 CCAGACCTTTCAGAGCAGCATGACCAAGATCTGGAGCGCTTCCGCGCGCGCAACCC 1200  
Db 1141 CCAGACCTTTCAGAGCAGCATGACCAAGATCTGGAGCGCTTCCGCGCGCGCAACCC 1200  
Qy 1201 GAGATCGTGTACTACAGTATCATGGAGCAGCTGACGTGGGCGAGCAGCTGGAGATCGGC 1260  
Db 1201 GAGATCGTGTACTACCA- - - - -GGCCCGCTTACGTTGGGCGAGCAGCTGGAGATCGGC 1260  
Qy 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAGCAGCTGCTGCGCTGGGGCTTCAACACC 1320  
Db 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAGCAGCTGCTGCGCTGGGGCTTCAACACC 1320  
Qy 1321 CCGGCAAGAGCAGCAGAGAGGAGCGCCCTTCTGTTGGATGGGCTAGAGCTGCACGCC 1380  
Db 1321 CCGGCAAGAGCAGCAGAGAGGAGCGCCCTTCTGTTGGATGGGCTAGAGCTGCACGCC 1380  
Qy 1381 GACAAGTGGAGCTGCGAGCGCATCGAGCTGCGCGAGAGGAGAGCTGAGCGCTGAACGAC 1440  
Db 1381 GACAAGTGGAGCTGCGAGCGCATCGAGCTGCGCGAGAGGAGAGCTGAGCGCTGAACGAC 1440  
Qy 1441 ATCCAGAGCTGTGGGCAAGTGAATCTGGGCGAGCGAGATCTACCCCGGATCAAGGTG 1500  
Db 1441 ATCCAGAGCTGTGGGCAAGTGAATCTGGGCGAGCGAGATCTACCCCGGATCAAGGTG 1500  
Qy 1501 GCGCAGCTGTGAGCTGCGCGCGCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
Db 1501 GCGCAGCTGTGAGCTGCGCGCGCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
Qy 1561 GAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGGAGCGCGCTGACGGC 1620  
Db 1561 GAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGGAGCGCGCTGACGGC 1620  
Qy 1621 GTGTACTACGACCCCAAGGAGCTGTGGTGGCGAGATTCAGAGAGGAGGCGCGACGACG 1680  
Db 1621 GTGTACTACGACCCCAAGGAGCTGTGGTGGCGAGATTCAGAGAGGAGGCGCGACGACG 1680

```
Db 1615 GTGTACTACGACCCAGCAGGAGACCTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAG 1674
Qy 1681 TGGACCTTACAGATCTTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1740
Db 1675 TGGACCTTACAGATCTTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1734
Qy 1741 ATGGGACCGCCACACCAAGCAGCTGAAGCAGCTGACCCAGAGCCGCTGCAAGAGATCGCC 1800
Db 1735 ATGGGACCGCCACACCAAGCAGCTGAAGCAGCTGACCCAGAGCCGCTGCAAGAGATCGCC 1794
Qy 1801 ATGGAGAGCATCGTGTACTTGGGGCAAGACCCCAAGTTCCGCTCGCCATCCAGAAAGGAG 1860
Db 1795 ATGGAGAGCATCGTGTACTTGGGGCAAGACCCCAAGTTCCGCTCGCCATCCAGAAAGGAG 1854
Qy 1861 ACCTGGGAGACCTGGTGGACCGACTACTGCGAGCCCACTTGGATCCCGAGTGGGAGTTC 1920
Db 1855 ACCTGGGAGACCTGGTGGACCGACTACTGCGAGCCCACTTGGATCCCGAGTGGGAGTTC 1914
Qy 1921 GTCAACACCCCCCTTACGTGGACGGCGCCCAACCGCGAGACCAGATCGGCAAGCGCGGC 2040
Db 1915 GTGAAACACCCCCCTTACGTGGACGGCGCCCAACCGCGAGACCAGATCGGCAAGCGCGGC 2034
Qy 2041 TAGCTGACCGACCGGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAG 2100
Db 2035 TAGCTGACCGACCGGGCGCGGCAAGATCGTGAGCTGACCGAGACCAACCAAGAG 2094
Qy 2101 ACCGAGCTGACGGCATCCAGCTGGCCCTGACGAGACCGGACCGGAGTGAACATCGTG 2160
Db 2095 ACCGAGCTGACGGCATCCAGCTGGCCCTGACGAGACCGGAGTGAACATCGTG 2154
Qy 2161 ACCGACAGCGCATACGCCCTTGGGCGATCATCCAGGCCCGACCGACAGAGCGAGCGAG 2220
Db 2155 ACCGACAGCGCATACGCCCTTGGGCGATCATCCAGGCCCGACCGACAGAGCGGAG 2214
Qy 2221 CTGGTGAAACAGATCATCGAGAGCTGTATCAAGAGAGAGAGTGTACCTGAGCTGGGTG 2280
Db 2215 CTGGTGAAACAGATCATCGAGAGCTGTATCAAGAGAGAGAGTGTACCTGAGCTGGGTG 2274
Qy 2281 CCCGCCACAAAGGCGCATCGCGGCAAGCAGATCGACAAGCTGGTGACCAAGGGGCATC 2340
Db 2275 CCCGCCACAAAGGCGCATCGCGGCAAGCAGATCGACAAGCTGGTGACCAAGGGGCATC 2334
Qy 2341 CGCAAGGTGCTGTTCTTGGACGGCGATCGATGGGGCATCGTGATCTACCAATACATGGAC 2400
Db 2335 CGCAAGGTGCTGTTCTTGGACGGCGATCGATGGGGCATCGTGATCTACCAATACATGGAC 2394
Qy 2401 GACCTGTACCTGGGACGCGCGCCCTAGAGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Db 2395 GACCTGTACCTGGGACGCGCGCCCTAGAGATCGATTAAAGCTTCCCGGGCTAGCAC 2454
Qy 2461 GGTGAATTC 2469
Db 2455 GGTGAATTC 2463
```

```
RESULT 3
AX455916
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 32 from Patent W00204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE
1
AUTHORS zur Megede,J., Barnett,S.W., Engelbrecht,S. and van Rensburg,E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides.
```

```
JOURNAL polypeptides and uses thereof
Patent: WO 0204493-A 32 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source Location/Qualifiers
1..2457
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YMMW"
```

## ORIGIN

```
Query Match 97.8%; Score 2415.4; DB 6; Length 2457;
Best Local Similarity 99.3%; Pred. No. 1e-243;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

Qy 1 GTCAGCGCCACCATGCGCGAGGCCATGAGCGAGCCACAGCGCCCAACATCTCTGATGCAG 60
Db 1 GTCAGCGCCACCATGCGCGAGGCCATGAGCGAGCCACAGCGCCCAACATCTCTGATGCAG 60
Qy 61 CGCAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
Db 61 CGCAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
Qy 121 CACATCGCCCGCAACTTCCCGCGCCCGCCCGCAAGAGGGCTGTGGAAAGTGCAGCAAGAG 180
Db 121 CACATCGCCCGCAACTTCCCGCGCCCGCCCGCAAGAGGGCTGTGGAAAGTGCAGCAAGAG 180
Qy 181 GGCACACAGATGAAGACTGCAACGAGCGCGCAGCCCACTTCTTCGCGAGGACCTGGCC 240
Db 181 GGCACACAGATGAAGACTGCAACGAGCGCGCAGCCCACTTCTTCGCGAGGACCTGGCC 240
Qy 241 TTCCCGCCAGGGCAAGCGCGGAGTTCCCGCAGGAGCAGAAACCGCGCAACAGCCCCACC 300
Db 241 TTCCCGCCAGGGCAAGCGCGGAGTTCCCGCAGGAGCAGAAACCGCGCAACAGCCCCACC 300
Qy 301 AGCGCGAGCTGCAAGTGTGCGCGCGACAACCCCCCGCAGCGAGCCCGCGCGCGCGCAG 360
Db 301 AGCGCGAGCTGCAAGTGTGCGCGCGACAACCCCCCGCAGCGAGCCCGCGCGCGCGCAG 360
Qy 361 GGCACCTTGAACCTTCCCGCAGATCACCTGTGTGCGAGGCGCCCTGTGAGCATCAAGTG 420
Db 361 GGCACCTTGAACCTTCCCGCAGATCACCTGTGTGCGAGGCGCCCTGTGAGCATCAAGTG 420
Qy 421 GCGGCGCAGATCAAGGAGGCGCTGTGTGGACAACCGCGCGCGCAGACACCGTGTGGAGGAG 480
Db 421 GCGGCGCAGATCAAGGAGGCGCTGTGTGGACAACCGCGCGCGCAGACACCGTGTGGAGGAG 480
Qy 481 ATGAGCCTTCCCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGCGCTTCATCAAG 540
Db 481 ATGAGCCTTCCCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGCGCTTCATCAAG 540
Qy 541 GTGCGCGAGTACCAACAGATCTCTGATGAGATCTGCGGCAAGAGGCGCATCGGCACCGTG 600
Db 541 GTGCGCGAGTACCAACAGATCTCTGATGAGATCTGCGGCAAGAGGCGCATCGGCACCGTG 600
Qy 601 CTGATCGGCGCCACCCCGCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCGCCACCCCGCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAACCTTCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCGATG 720
Db 661 ACCCTGAACCTTCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCGATG 720
Qy 721 GACGCGCCCAAGGTGAAGCTGGCCCTTGAACCGAGGAGAGATCAAGGCGCTTACCGCC 780
Db 721 GACGCGCCCAAGGTGAAGCTGGCCCTTGAACCGAGGAGAGATCAAGGCGCTTACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAGAGGAGGCAAGATCAACAGATCGCGCCCGCAAGACCCCTAC 840
Db 781 ATCTGCGAGGAGATGGAGAGGAGGCAAGATCAACAGATCGCGCCCGCAAGACCCCTAC 840
Qy 841 AACACCCCGCTGTTCGCCATCAAGAAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCGCTGTTCGCCATCAAGAAAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
```

841 AACACCCCGCTGTTCCGCATCAAGAGAAAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
901 TTCCGCGAGCTGAACAACAGCGACCCAGGACTTCTGGAGGTGCGACTGGGCATCCCCCAC 960  
902 TTCCGCGAGCTGAACAACAGCGACCCAGGACTTCTGGAGGTGCGACTGGGCATCCCCCAC 960  
961 CCAGCGGCTGAGAGAGAGAGCGTGCCTGCTGGAGGTGGCGAGCGCTACTTTC 1020  
962 CCAGCGGCTGAGAGAGAGAGCGTGCCTGCTGGAGGTGGCGAGCGCTACTTTC 1020  
1021 AGCGTGCCTTGGAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
1022 AGCGTGCCTTGGAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080  
1081 AACGAGACCCCGGCTACCGTACAGTACAACTGCTGCCCGGAGGTGGAGGGCAGC 1140  
1082 AACGAGACCCCGGCTACCGTACAGTACAACTGCTGCCCGGAGGTGGAGGGCAGC 1140  
1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCCCGCAACCCC 1200  
1142 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCCCGCAACCCC 1200  
1201 GAGATCGTGTACTACCAAGTACATGAGACCTGTAGTGGCGAGCCTCGAGATCGGC 1260  
1202 GAGATCGTGTACTACCAAGTACATGAGACCTGTAGTGGCGAGCCTCGAGATCGGC 1260  
1203 GAGATCGTGTACTACCAAGTACATGAGACCTGTAGTGGCGAGCCTCGAGATCGGC 1260  
1261 CAGCAGCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTCCACACC 1320  
1262 CAGCAGCGCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTCCACACC 1320  
1263 CAGCAGCGCGCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGGCTTCCACACC 1320  
1321 CCCAGCAAGAGCAGCAGAGAGCGCCCTTCTGTGGATGGGCTAGAGCTGCACGCC 1380  
1322 CCCAGCAAGAGCAGCAGAGAGCGCCCTTCTGTGGATGGGCTAGAGCTGCACGCC 1380  
1323 CCCAGCAAGAGCAGCAGAGAGCGCCCTTCTGTGGATGGGCTAGAGCTGCACGCC 1380  
1324 CCCAGCAAGAGCAGCAGAGAGCGCCCTTCTGTGGATGGGCTAGAGCTGCACGCC 1380  
1325 GACAAGTGGACCGTGGAGCGCCATGAGCTGCGCGAGAGGAGCTGACCGTGAACGAC 1440  
1326 GACAAGTGGACCGTGGAGCGCCATGAGCTGCGCGAGAGGAGCTGACCGTGAACGAC 1440  
1369 GACAAGTGGACCGTGGAGCGCCATGAGCTGCGCGAGAGGAGCTGACCGTGAACGAC 1428  
1441 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1500  
1442 ATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1488  
1501 CGCAGCTGTGCAAGCTGTGCGGCGCGCAAGCGCTGACCGCATCTGCGCCCTGACC 1560  
1489 CGCAGCTGTGCAAGCTGTGCGGCGCGCAAGCGCTGACCGCATCTGCGCCCTGACC 1548  
1561 GAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCGCGTGCACGGC 1620  
1549 GAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCGCGTGCACGGC 1608  
1621 GTGTACTACGACCCAGCAAGGACTGTGGTGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1680  
1609 GTGTACTACGACCCAGCAAGGACTGTGGTGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1668  
1681 TGGACCTTACAGATCTTACAGGAGCGCTTCAAGAACCTGGAAGCGGCAAGTACGCCAAG 1740  
1669 TGGACCTTACAGATCTTACAGGAGCGCTTCAAGAACCTGGAAGCGGCAAGTACGCCAAG 1728  
1741 ATGCGCACCGCCACACCGAGCTGAGCAGCTGACCGAGCGCTGCGAGAGATCGCC 1800  
1729 ATGCGCACCGCCACACCGAGCTGAGCAGCTGACCGAGCGCTGCGAGAGATCGCC 1788  
1801 ATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTTCCGAGAGGAG 1860  
1789 ATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGAGAGGAG 1848  
1861 ACCTGGAGACCTGGTGGAGCCGATCTGCGAGCGCAGCTGAGTCCCGAGTGGAGTTC 1920  
1849 ACCTGGAGACCTGGTGGAGCCGATCTGCGAGCGCAGCTGAGTCCCGAGTGGAGTTC 1908  
1921 GTGAGACACCCCGCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC 1980  
1909 GTGAGACACCCCGCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC 1968

QY 1981 GCCGAGACCTTCTTACGTGGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGC 2040  
DB 1969 GCCGAGACCTTCTTACGTGGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGC 2028  
QY 2041 TACGTGACCGACCGGCGCGCAGAGATCTGTAGCGCTGAGCGGAGCACCACCAAGCAGAG 2100  
DB 2029 TACGTGACCGACCGGCGCGCAGAGATCTGTAGCGCTGAGCGGAGCACCACCAAGCAGAG 2088  
QY 2101 ACCGAGCTGACGAGGCGCATCCAGCTGGCGCCCTGAGGACAGCGCGCAGCGAGTGAACATCGTG 2160  
DB 2089 ACCGAGCTGACGAGGCGCATCCAGCTGGCGCCCTGAGGACAGCGCGCAGCGAGTGAACATCGTG 2148  
QY 2161 ACCGAGCGCAGTACGCGCTTGGCGCATCATCAGCGCCAGCGCAGCGAGCGAGCGAG 2220  
DB 2149 ACCGAGCGCAGTACGCGCTTGGCGCATCATCAGCGCCAGCGCAGCGAGCGAGCGAG 2208  
QY 2221 CTGCTGACCGACAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGGTG 2280  
DB 2209 CTGCTGACCGACAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGGTG 2268  
QY 2281 CCCGCGCAGAGGCGCATCGCGCGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATC 2340  
DB 2269 CCCGCGCAGAGGCGCATCGCGCGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATC 2328  
QY 2341 CGCAAGGTGCTTCTTGGAGCGGCGCATCGATCGCGCGCATCTGTGATCTACGATACATGAC 2400  
DB 2329 CGCAAGGTGCTTCTTGGAGCGGCGCATCGATCGCGCGCATCTGTGATCTACGATACATGAC 2388  
QY 2401 GACCTGTACGTGGCGCGCGCGCTAGGATCGATTAAGAGCTTCCCGGGCTAGCACC 2460  
DB 2389 GACCTGTACGTGGCGCGCGCGCTAGGATCGATTAAGAGCTTCCCGGGCTAGCACC 2448  
QY 2461 GGTGAATTC 2469  
DB 2449 GGTGAATTC 2457

RESULT 4  
BD263706  
LOCUS  
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.  
ACCESSION BD263706  
VERSION BD263706.1 GI:33073474  
KEYWORDS JP 2002533124-A/73.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 2312)  
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: JP 2002533124-A 73 08-OCT-2002;  
COMMENT CHIRON CORP  
OS Artificial Sequence  
PN JP 2002533124-A/73  
PD 08-OCT-2002  
PF 30-DEC-1999 JP 2000591193  
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI  
KARIN HARTOG,  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC  
CI2N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
CI2N5/10,  
PC CI2N7/00, CI2P21/02, CI2N15/00, CI2N5/00, A61K37/02 CC  
Description of Artificial Sequence: FS(-).protmod.Rtopt(+) PH Key  
FT source 1..2312 /organism='Artificial Sequence'.  
FT Location/Qualifiers  
1..2312

		/organism="synthetic construct" /mol_type="genomic DNA" /db_xref="taxon:32630"					
ORIGIN							
Query Match		83.1%; Score 2052; DB 6; Length 2312;					
Best Local Similarity		93.8%; Pred. No. 9.6e-205;					
Matches 2165; Conservative		0; Mismatches 136; Indels 12; Gaps 2;					
QY	170	GCGGCAAGGAGGCCACACAGATGAAGGATGCAACCGAGCGCCAGGCCAACTTCTTCGGCG	229	QY	1118	TGCCCCAGGGCTGGAAAGGCGAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG	1177
DB	1	GCGGCGCGAAGGACACCAATGAAGATTGCATCGAGAGACAGGCTTAATTTCTTCGGCG	60	DB	961	TGCCCCAGGGCTGGAAAGGCGAGCCCCGCATCTTCCAGAGCAGCATGACCAAGATCCTGG	1020
QY	230	AGGACTTGGCTTCCCGCCAGGGAAGCGCCGAGTTCGCCAGCGAGCAGAAACCGCGCCA	289	QY	1178	AGCCCTTCGGCGCCCGCAACCCCGAGATCGTGATCTTACAGTACATGGAGCAGCTGTACG	1237
DB	61	AGGACTTGGCTTCCCGCCAGGGAAGCGCCGCGAGTTTCAGCAGCGAGCAGAACCCGCGCCA	120	DB	1021	AGCCCTTCGGCAAGCAGAAACCCCGCATCTGTGATCTACCAAGTACATGGAGCAGCTGTACG	1080
QY	290	ACAGCCCCACACCGCGAGCTGCAAGTTCGCGGGG-----ACAACCCCGCGAGCGAGG	343	QY	1238	TGGGAGCGACCTTGGAGATCGGCCAGCACCGGCCCAAGATCGAGGAGCTGCGCAAGCACC	1297
DB	121	ACAGCCCCACCGCGCGAGCTGCAAGTTCGCGGGGCGCGAGAAACAACAGCCTTGAGCGAGG	180	DB	1081	TGGGAGCGACCTTGGAGATCGGCCAGCACCGCCACCAAGATCGAGGAGCTGCGCGCAGCAC	1140
QY	344	CGGGCGCGAGCGCCAGGGCACCTG-----AATTTCCCGCAGATCACCTGTGGCAGC	397	QY	1298	TGCTGGCTGGGGCTTCAACACCCCGCAACAAGAACCAAGAGAGCCCCCTTCTCTGT	1357
DB	181	CGGGCGCGAGCGCCAGGGCACCTGTGAGCTTCAACTTCCCGCAGATCACCTGTGGCAGC	240	DB	1141	TGCTGGCTGGGGCTTCAACACCCCGCAACAAGAACCAAGAGAGCCCCCTTCTCTGT	1200
QY	398	GCCCCCTGTGTAGCATCAAGGTGGCGGCCAGATCAAGGAGGCGCTGTCTGGACACCGGCG	457	QY	1358	GGATGGCTTACGAGCTGCAACCCCGCAAGTGGACCGTGGCAAGCTGAACTGGCGCCAGGA	1417
DB	241	GCCCCCTGTGTACCATCAGATCGGCGGCCAGCTCAAGAGGCGCTGTCTGGACACCGGCG	300	DB	1201	GGATGGCTTACGAGCTGCAACCCCGCAAGTGGACCGTGGCAAGCTGAACTGGCGCCAGGA	1260
QY	458	CGACGACACCGTGTGGAGGAGATGAGCTGCGCGCAAGTGGAGCGCCCAAGATGATCG	517	QY	1418	AGGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACTGGCGCCAGCC	1477
DB	301	CGACGACACCGTGTGGAGGAGATGAACCTGCGCGCAAGTGGAGCGCCCAAGATGATCG	360	DB	1261	AGGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGGTGGCAAGCTGAACTGGCGCCAGCC	1320
QY	518	CGGGCATCGCGGCTTCATCAAGGTGGCGGCCAGTACGACCAAGATCTGTGATCTGGCG	577	QY	1478	AGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCC	1537
DB	361	CGGGCATCGGGGGCTTCATCAAGGTGGCGGCCAGTACGACCAAGATCTGTGATCTGGCG	420	DB	1321	AGATCTACCGCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGGCACCAAGGCC	1380
QY	578	GCAAGAAGGCGCATCGGCACCGTCTGATCGGCCCGCCACCCCGTGAACATCATCGGCGCA	637	QY	1538	TGACCAACATCTGTGCCCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGA	1597
DB	421	GCCACAAGGCCATTCGGCACCGTCTGTGTGGCGGCCACCCCGTGAACATCATCGGCGCA	480	DB	1381	TGACCGAGGTGATTCCTCCCTTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGA	1440
QY	638	ACATGTGTACCCAGCTGGGCTGCACCTGAACTTCCCGCATCAGCCCATCAGACCGGTGC	697	QY	1598	TCCTGCGCAGCGCCGTGCAAGGTGCTGTAACAACCCAGCAAGGAGCTGTGCGCGCCAGGA	1657
DB	481	ACCTGTGTACCCAGATCGGCTGCACCTGAACTTCCCGCATCAGCCCATCAGACCGGTGC	540	DB	1441	TCCTGAAGAGAGCCCGTGTGCAAGGTGTAACAACCCAGCAAGGAGCTGTGCGCGCCAGGA	1500
QY	698	CGGTGAAGCTGAAGCCCGGATGGAACCGGCCCAAGGTGAAGAGTGGCGCCCTGACCGAGG	757	QY	1658	TCAGAAGCAGGGCCACGACCAAGTGGACCTTACAGATCTTACAGAGCCCTTCAAGAAC	1717
DB	541	CGGTGAAGCTGAAGCCCGGGAATGGAACCGGCCCAAGGTGAAGAGTGGCGCCCTGACCGAGG	600	DB	1501	TCAGAAGCAGGGCCACGAGCCAGTGGACCTTACAGATCTTACAGGAGCCCTTCAAGAAC	1560
QY	758	AGAAGATCAAGGCGCTTACCGCATCTGCGAGGAGATGGAGAAGGAGGCGCAAGATCACCA	817	QY	1718	TGAAGACCGGCAAGTACGCCAAGATCGGCACCGGCCACCAACAGAGCTGAAGCAGCTGA	1777
DB	601	AGAAGATCAAGGCGCTTGTGGAGATCTGCACCGAGATGGAGAAGGAGGCGCAAGATCAGCA	660	DB	1561	TGAAGACCGGCAAGTACGCCCGCATGCGCGCGGCCACCAACAGCAGCTGAAGCAGCTGA	1620
QY	818	AGATCGGCCCCGAGAACCCCTTACAACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCA	877	QY	1778	CCGAGCCGTGCAAGAGATCGCCATCGAGAGATCTGTGATCTGGGGCAAGACCCCCCAAGT	1837
DB	661	AGATCGGCCCCGAGAACCCCTTACAACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCA	720	DB	1621	CCGAGCCGTGCAAGAGGTGAGCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGT	1680
QY	878	CAAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG	937	QY	1838	TCGCGCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTTGGGAGGCCA	1897
DB	721	CAAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG	780	DB	1681	TCAAGCTGCCCATCCAGAAGGAGACCTGGGAGGCGCTGGTGGATGGAGTACTTGGCAGGCCA	1740
QY	938	AGGTGCAAGTGGGCATCCCGCACCCCGCGGCTTGAAGAAGAAGAGCGGTGACCGGTGC	997	QY	1898	CTTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTGGTGAAGCTGTGTGTTACCGC	1957
DB	781	AGGTGCAAGTGGGCATCCCGCACCCCGCGGCTTGAAGAAGAAGAGCGGTGACCGGTGC	840	DB	1741	CCTTGGATCCCGAGTGGGAGTTCTGTAACAACCCCGCCCTGGTGAAGCTGTGTGTTACCGC	1800
QY	998	TGAGCGTGGCGAGCGCTACTTACGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG	1057	QY	1958	TGAGAAGGAGAGCCCATCTATCGCGCGCGAGACCTTCTACGTGGAGCGCGCCCGCAACCGCG	2017
DB	841	TGAGCGTGGCGAGCGCTACTTACGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG	900	DB	1801	TGAGAAGGAGAGCCCATCTGTGGGCGCGAGACCTTCTACGTGGAGCGCGCCCAACCGCG	1860
QY	1058	CCTTACCATCCCGAGCATCAACAAGACACCCCGGCATCCGTTACAGTACAAAGTGC	1117	QY	2018	AGACCAAGATTCGGCAAGGCGGCTACGTGACCGACCGGGGCGCGGAGAGATCGTGAGCC	2077
DB	901	CCTTACCATCCCGAGCATCAACAAGACACCCCGGCATCCGTTACAGTACAAAGTGC	960	DB	1861	AGACCAAGCTGGGCAAGGCGGCTACGTGACCGACCGGGGCGCGGAGAGGTGTGTAGCA	1920
QY				QY	2078	TGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTCGAGGACA	2137
DB				DB	1921	TCGCCGACACCAACCAAGAGACCGAGCTGCAGGCCATCCACCTGSCCTCGCAGGACA	1980
QY				QY	2138	GCGGCAAGCGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTTGGGTCATCTCAGGCC	2197
DB				DB	1981	GCGGCTTGGAGGTGAACATCTGTGACCGCAGCAGCTACGCCCTTGGGTCATCTCAGGCC	2040
QY				QY	2198	AGCTCCACAGAGCGAGCGAGCTGGTGAAACCAATCATCGAGCAGCTGATCAAGAAGG	2257

```
Db 2041 AGCCCGCAAGAGCGAGCGAGCTGGTGAACCCAGATCATCGACGCTGATCAAGAAGG 2100
Qy 2258 AGAAGGTGTACTTCTGAGCTGGTGCCTGCGCCCAACAAGGGCATCGCGCGCAACAGCAGATCG 2317
Db 2101 AGAAGGTGTACTTCTGAGCTGGTGCCTGCGCCCAACAAGGGCATCGCGCGCAACAGCAGTGG 2160
Qy 2318 ACAAGCTGGTGAAGCAAGGCGCATCCGCAAGGTGCTGTTCTTGGAGCGGATCGATCGCGGCA 2377
Db 2161 ACAAGCTGGTGAAGCGCGCGCATCCGCAAGGTGCTGTTCTTGGAGCGGATCGATCGCGGCA 2220
Qy 2378 TCGTGATCTACCACTACATGAGCAACCTGTACGTGGCGAGCGCGGCGCTAGGATCGATT 2437
Db 2221 TCGTGATCTACCACTACATGAGCAACCTGTACGTGGCGAGCGCGGCGCTAGGATCGATT 2280
Qy 2438 AAAAGCTTCCCGGGGCTAGCACCGGTGAATT 2469
Db 2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATT 2312

RESULT 5
LOCUS CQ870576 2312 bp DNA linear PAT 13-SEP-2004
DEFINITION Sequence 84 from Patent EP1433851.
ACCESSION CQ870576
VERSION CQ870576.1 GI:52000092
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
1. Barnett, S., Zumegede, J., Srivastava, I., Lian, Y., Hartog, K.,
AUTHORS Liu, H., Greer, C., Selby, M. and Walker, C.
TITLE Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL Patent: EP 1433851-A 84 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES
source
1. .2312
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="Description of Artificial Sequence:
FS(-).protmod.Rtopt(+)"
ORIGIN

Query Match 83.1%; Score 2052; DB 6; Length 2312;
Best Local Similarity 93.6%; Pred. No. 9.6e-206;
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

Qy 170 GCGGCAAGGAGGGCCACCAGATGAAGACTGCAACGAGCGCCAGGCCAACTTCTTCGCG 229
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGCAGGCTAATTCTTCGCG 60
Qy 230 AGAAGCTGGCTTCCCGCAGGCGAAGCGCGCGAGTTCCCGAGGAGCAGAAACCGGCCA 289
Db 61 AGAAGCTGGCTTCCCGCAGGCGAAGCGCGCGAGTTTCCAGCAGCAGCAGCAGCAGCGCCA 120
Qy 290 ACAGCCCCACAGCGCGAGCTGAGGTGCGCGCG-----ACAACCCCGCAGCGAGG 343
Db 121 ACAGCCCCACCGCGCGAGCTGAGGTGTTGGGCGCGGAGAAACAACAGCTGAGCGAG 180
Qy 344 CCGCGCGCAGCGCGCAGCGCACCTG-----AACTTCCCGCCAGATCACCTGTGGCAGC 397
Db 181 CCGCGCGCAGCGCGCAGCGCACCTGAGCTTCACTTCCCGCCAGATCACCTGTGGCAGC 240
Qy 398 GCGGCTGGTGAAGATCAAGGTGGCGCGCAGATCAAGAGAGCGCTCTGAGCAACCGGCG 457
Db 241 GCGGCTGGTGAAGATCAAGGTGGCGCGCAGCTCAAGAGAGCGCTCTGAGCAACCGGCG 300
Qy 458 CCGAGCAGCAGCTGGTGGAGAGATGAGCTGCGCGCAAGTGGAGCGCCCAAGATGATCG 517
Db 301 CCGAGCAGCAGCTGGTGGAGAGATGAGCTGCGCGCAAGTGGAGCGCCCAAGATGATCG 360
```

```
Qy 518 GCGGCATCGCGCGCTTCAATCAAGGTGCGCCAGTACGACCAAGATCCTGATCAGATCTGCG 577
Db 361 GCGGCATCGCGCGCTTCAATCAAGGTGCGCGAGTACGACCAAGATCCTGATCAGATCTGCG 420
Qy 578 GCAAGAGGCGCATCGGCACCGTCTGATCGGCCCGCCCGCGTGAACATCATCGCGCGCA 637
Db 421 GCCAAGGCGCATCGGCACCGTCTGATCGGCCCGCCCGCGTGAACATCATCGCGCGCA 480
Qy 638 ACATGCTGACCGAGCTGGGCTGCAACCTGAACTTCCCATCAGCCCCCATCGAGACCGTGC 697
Db 481 ACCTGCTGACCGAGATCGGCTGCACCTGAACTTCCCATCAGCCCCCATCGAGACCGTGC 540
Qy 698 CCGTGAAGCTGAAGCGCGCGCATGGAACCGCCCAAGGTGAAGCACTGCGCCCCCTGACCGAG 757
Db 541 CCGTGAAGCTGAAGCGCGCGCATGGAACCGCCCAAGGTGAAGCACTGCGCCCCCTGACCGAG 600
Qy 758 AGAAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAAGAGGCGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCGCTGCTGGAGATGTCACCGAGATGGAAGAGGCGCAAGATCACCA 660
Qy 818 AGATCGGCGCGCGAAGACCCCTTACAAACACCCCGCTGTTCCGCATCAAGAAGAGCAGCA 877
Db 661 AGATCGGCGCGCGAAGACCCCTTACAAACACCCCGCTGTTCCGCATCAAGAAGAGCAGCA 720
Qy 878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGG 937
Db 721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGG 780
Qy 938 AGTGCAGCTGGGCGATCCCGCACCCCGCGCGCTGGAAGAGAGGCGTGCCTGTC 997
Db 781 AGTGCAGCTGGGCGATCCCGCACCCCGCGCGCTGGAAGAGAGGCGTGCCTGTC 840
Qy 998 TGCAGCTGGGCGCGCGCTTCTTCAAGCTGCGCTTGCAGCAGGACTTCCGCAAGTACACCG 1057
Db 841 TGCAGCTGGGCGCGCGCTTCTTCAAGCTGCGCTTGCAGCAGGACTTCCGCAAGTACACCG 900
Qy 1058 CTTTACCATCCCGCAGCATCAACAAGAGACCCCGCGCATCCGCTACAGTACAACTGTC 1117
Db 901 CTTTACCATCCCGCAGCATCAACAAGAGACCCCGCGCATCCGCTACAGTACAACTGTC 960
Qy 1118 TGCAGCTGGGCGCGCGCGCGCGCGCGCGCTGGAAGAGAGGCGTGCCTGTC 1177
Db 961 TGCAGCTGGGCGCGCGCGCGCGCGCGCGCGCGCATCCGCTACAGTACAACTGTC 1020
Qy 1178 AGCCTTTCGCGCGCGCAACCCCGAGATCGTGATCTACCACTACAGTACAGACCTGTACG 1237
Db 1021 AGCCTTTCGCGCGCGCAACCCCGAGATCGTGATCTACCACTACAGTACAGACCTGTACG 1080
Qy 1238 TGGGCGAGCGACCTGGAGATCGGCGCAGCACCGCGCCAAAGATCGAGGAGCTGCGCAAGCAC 1297
Db 1081 TGGGCGAGCGACCTGGAGATCGGCGCAGCACCGCGCCAAAGATCGAGGAGCTGCGCGCAGCAC 1140
Qy 1298 TGTGCTGGCTGGGCGTTTCAACCCCGCAAGAAGCAGCAGAGGAGCGCGCTTCTGT 1357
Db 1141 TGTGCTGGCTGGGCGTTTCAACCCCGCAAGAAGCAGCAGAGGAGCGCGCTTCTGT 1200
Qy 1358 GGATGGCTACGAGCTGCGACCCGCAAGTGGACCGTGCAGCGCCCATCGAGCTGCGCGCAGGA 1417
Db 1201 GGATGGCTACGAGCTGCGACCCCGCAAGTGGACCGTGCAGCGCCCATCGAGCTGCGCGCAGGA 1260
Qy 1418 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCC 1477
Db 1261 AGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCC 1320
Qy 1478 AGATCTACCCCGCGCATCAAGTGGCGCGCGCTGTGCAAGCTGCTGCGCGCGCGCGCAGGCC 1537
Db 1321 AGATCTACCCCGCGCATCAAGTGGCGCGCGCTGTGCAAGCTGCTGCGCGCGCGCGCAGGCC 1380
Qy 1538 TGACCGCAGCATCTGCGCGCTTACCGAGGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1597
Db 1381 TGACCGGAGTGTATCCCTTACCGAGGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1440
```



1598 TCCTGCGCAGCCGTCACCGGCTGTACTACACCCCAAGGACCTGGTGGCGAGA 1657  
 1441 TCCTGAAGAGCCCGTGCAGAGTGTACTACGCCCAAGGACCTGGTGGCGAGA 1500  
 1658 TCAGAAAGAGGCGCAGCAGCACTGAGCCTTACAGATCTACAGAGGCGCTTCAAGAAC 1717  
 1501 TCAGAAAGAGGCGCAGGCGAGTGGACCTTACAGATCTACAGAGGCGCTTCAAGAAC 1560  
 1718 TGAAGACCGCAGTACGCAAGATGGCAAGCCCGCCACCAACCAAGCAGTGAAGCAGCTGA 1777  
 1561 TGAAGACCGCAGTACGCAAGATGGCAAGCCCGCCACCAACCAAGCAGTGAAGCAGCTGA 1620  
 1778 CCAGAGCCGTCAGAAAGATCCCATCGAGAGCATCGTGATCTGGGGCAAGACCCCAAGT 1837  
 1621 CCAGAGCCGTCAGAAAGTGAACACCGAGAGCATCGTGATCTGGGGCAAGATCCCAAGT 1680  
 1838 TCAGCCCTGCCCATTCAAGAAAGAGACCTTGGAGACCTGGTGGAGCCGACTACTGGCAGGCCA 1997  
 1681 TCAAGCTGCCCATTCAAGAAAGAGACCTTGGAGAGCCTGGTGGAGTACTGGCAGGCCA 1740  
 1898 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCTGGAAGCTGTGGTACCAAGC 1957  
 1741 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCTGGAAGCTGTGGTACCAAGC 1800  
 1958 TGGAGAAAGAGCCCATCATCGCGCCGAGACCTTTACGTGGACGGCGCCCAACCGCG 2017  
 1801 TGGAGAAAGAGCCCATCGTGGCGCCGAGACCTTTACGTGGACGGCGCCCAACCGCG 1860  
 2018 AGAACCAAGATCGGCAAGCGCGCTTACGTGACCAACCGCGCCGCGCAGAGATCGTGAGCC 2077  
 1861 AGAACCAAGTGGGCAAGCGCGCTTACGTGACCAACCGCGCCGCGCAGAGTGGTGAGCA 1920  
 2078 TGAACCGAGACCAACCAAGAGACCGAGCTGCGAGCCATCCAGCTGGCCCTGCAGGACA 2137  
 1921 TCGCCGACACCAACCAAGAGACCGAGCTGCGAGCCATCCAGCTGGCCCTGCAGGACA 1980  
 2138 GCGGCAGCAGAGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCGATCATCCAGGCC 2197  
 1981 GCGGCCTGGAGGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCGATCATCCAGGCC 2040  
 2198 AGCCCGACAAGAGCGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAGG 2257  
 2041 AGCCCGACAAGAGCGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAGG 2100  
 2258 AGAAGGTGTAACCTGAGCTGGGTCGCCGCCCAAGGCGATCGCGGCAAGCAGATCG 2317  
 2101 AGAAGGTGTAACCTGAGCTGGGTCGCCGCCCAAGGCGATCGCGGCAAGCAGATCG 2160  
 2318 ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGCA 2377  
 2161 ACAAGCTGGTGAGCGCGCATCCGCAAGGTGCTGTTCTTGAACGGCATCGATGGCGGCA 2220  
 2378 TCGTGATCTAACAGTATACAGCAGCTGTGACGTGGGACGGCGGCTTGGATGATTT 2437  
 2221 TCGTGATCTAACAGTATACAGCAGCTGTGACGTGGGACGGCGGCTTGGATGATTT 2280  
 2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2469  
 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 6  
 AR373389  
 LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003  
 DEFINITION Sequence 84 from patent US 6602705.  
 AR373389  
 AR373389  
 AR373389.1 GI:40075492  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 UNCLASSIFIED.  
 REFERENCE 1 (bases 1 to 2312)  
 AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.

TITLE Expression of HIV polypeptides and production of virus-like particles  
 JOURNAL Patent: US 6602705-A 84 05-AUG-2003;  
 FEATURES Location/Qualifiers  
 source 1. .2312  
 /organism="unknown"  
 /mol\_type="genomic DNA"  
 ORIGIN  
 Query Match 83.1%; Score 2052; DB 6; Length 2312;  
 Best Local Similarity 93.8%; Pred. No. 9,6e-206;  
 Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
 170 GCGGCAAGAGGCGCCACAGAGTGAAGGACTGACACCGAGGCGCCAGGCGCAATTTCTTCGCG 229  
 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60  
 230 AGACCTGCGCTTCCCGCAGGGCAAGCGCGGAGTTCCCGCAGCAGCAGAGCAACCGCGCA 289  
 61 AGACCTGCGCTTCTGCGAGGGCAAGCGCGGAGTTTCAGCAGCAGCAGCAGCAGCGCGCA 120  
 290 ACAGCCCCACAGCGCGGAGCTGCAGGTGTGGGGCGCGAGGCGGCGGCGGCGGCGG 343  
 121 ACAGCCCCACCGCGCGAGCTGCAGGTGTGGGGCGCGAGGCGGCGGCGGCGGCGG 180  
 344 CCGGCGCGCAGCGCGAGGCGACCCCTG-----AACTTCCCGCAGATCAACCTTGTGGCAGC 397  
 181 CCGGCGCGCAGCGCGAGGCGACCGTGCAGCTTCAACTTCCCGCAGATCAACCTTGTGGCAGC 240  
 398 GCCCCTGTGAGCATCAAGGTGGGCGCGCAGATCAAGAGGCGCTTGTGGACACCGCGCG 457  
 241 GCCCCTGTGAGCATCAGGATCGGCGCGCAGCTCAAGAGGCGCTGCTGCACACCGCGCG 300  
 458 CGACACACCGCTGCGAGGAGATGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 517  
 301 CGACACACCGCTGCGAGGAGATGAACTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
 518 GCGGCATCGCGCGCTTCAAGGTGCGCGAGTACGACAGATCTTGTGATCGAGATCTGCG 577  
 361 GCGGCATCGCGCGCTTCAAGGTGCGCGAGTACGACAGATCTTGTGATCGAGATCTGCG 420  
 578 GCAAGAGGCGCATCGGCGACCGTCTGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637  
 421 GCACAAAGGCGCATCGGCGACCGTCTGTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCA 480  
 638 ACATGCTGACCGAGTGGGCTGCACCTGAACTTCCCATCAGCGCGCGCGCGCGCGCGCG 697  
 481 ACCTGCTGACCGAGATCGGCTGCACCTGAACTTCCCATCAGCGCGCGCGCGCGCGCGCG 540  
 698 CCGTGAAGCTGAAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757  
 541 CCGTGAAGCTGAAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
 758 AGAAGATCAAGGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 817  
 601 AGAAGATCAAGGCGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 660  
 818 AGATCGCGCGCGAGAACCCCTCAACACCCCGCTGTTCCGCGCGCGCGCGCGCGCGCGCG 877  
 661 AGATCGCGCGCGAGAACCCCTCAACACCCCGCTGTTCCGCGCGCGCGCGCGCGCGCGCG 720  
 878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAGCGCGCGCGCGCGCGCGCGCG 937  
 721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAGCGCGCGCGCGCGCGCGCGCG 780  
 938 AGGTGAGCTGGGCGATCCCG 997  
 781 AGGTGAGCTGGGCGATCCCG 840  
 998 TGGACGTGGGCGAGCGCTACTTCCAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057  
 841 TGGACGTGGGCGAGCGCTACTTCCAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900



1058 CTTTACCATCCCGAGCATCAACACGAGAGCCCGGGATCCGCTACAGTACAACTGTC 1117  
Db CTTTACCATCCCGAGCATCAACACGAGAGCCCGGGATCCGCTACAGTACAACTGTC 960  
1118 TGCCCGAGGGCTGGAAGGGCAGCCCGAGCATCTTTCCAGAGCAGCATGACCAAGATCTCTGG 1177  
Db TGCCCGAGGGCTGGAAGGGCAGCCCGGGATCTTTCCAGAGCAGCATGACCAAGATCTCTGG 1020  
1178 AGCCCTTCCGGCCCGCAACCCCGAGATCGTGATCTTACCAAGTACATGGAACCTGTACG 1237  
Db AGCCCTTCCGGCCCGCAAGAACCCCGACATCGTGATCTTACCAAGTACATGGAACCTGTACG 1080  
1238 TGCGCAGGACCTGGAGATCGGCCAGCACCCCGCCAGATCGAGAGCTGGCGAAGCACC 1297  
Db TGCGCAGGACCTGGAGATCGGCCAGCACCCCGCCAGATCGAGAGCTGGCGCAGCACC 1140  
1298 TGCTCGCGCTGGGGCTTCAACACCCCGCAAGAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1417  
Db TGCTCGCGCTGGGGCTTCAACACCCCGCAAGAGTGGACCGTGCAGCCCATCGAGCTGCCGAGA 1260  
1418 AGGAGACTGGACCGTGAAACGACATCCAGAAAGTGGTGGGGCAAGCTGAACTGGGGCAGCC 1477  
Db AGGAGACTGGACCGTGAAACGACATCCAGAAAGTGGTGGGGCAAGCTGAACTGGGGCAGCC 1320  
1478 AGATCTACCCCGGATCAAGTGGCGCAGCTGTCAAGCTGTCTCGGGCGGCCCAAGGCC 1537  
Db AGATCTACCCCGGATCAAGTGGCGCAGCTGTCAAGCTGTCTCGGGCGGCCCAAGGCC 1380  
1538 TGACCGACATCGTGCCTTACCCGAGGAGGCGAGCTGGAGCTGGCGGAGAACCGCGAGA 1597  
Db TGACCGAGGTGATCCCTTACCCGAGGAGGCGAGCTGGAGCTGGCGGAGAACCGCGAGA 1440  
1598 TCCTGCGCGAGCCCGTGACCGCGGTGTACTACGACCCCGAGAGGACTGGTGGCCGAGA 1657  
Db TCCTGAGAGGAGCCCGTGACGAGGTGTACTACGACCCCGAGAGGACTGGTGGCCGAGA 1500  
1658 TCAGAGAGGAGGCGACGACGAGTGGACCTACCAAGTCTACGAGGAGCCCTTCAAGAAC 1717  
Db TCAGAGAGGAGGCGACGACGAGTGGACCTACCAAGTCTACGAGGAGCCCTTCAAGAAC 1560  
1718 TGAAGACCGGCAAGTACGCAAGTGGCGACCGCGCCACCAACAGCTGAAGCAGCTGA 1777  
Db TGAAGACCGGCAAGTACGCGCGCATGCGCGGCGCCACCAACAGCAGCTGAAGCAGCTGA 1620  
1778 CCGAGGCGGTGCAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGT 1837  
Db CCGAGGCGGTGCAAGATCGCAACCGAGAGCATCGTGATCTGGGGCAAGATCCCAAGT 1680  
1838 TCCGCTGCCCATTCCAGAGGAGACCTGGGAGACCTGGTGACCGGACTCTGGCAGGCCA 1897  
Db TCAAGCTGCCCCATCCAGAGGAGACCTGGGAGGCTGGTGATGAGTACTGGCAGGCCA 1740  
1898 CTTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGGCTGGTGAAGCTGTGTACCGAGC 1957  
Db CTTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGGCTGGTGAAGTACTGGTACCGAGC 1800  
1958 TGAGAAAGGAGCCCATCATCGGGCCCGAGACCTTCTTACGTGGAGCGCGCCCGCAACCGCG 2017  
Db TGAGAAAGGAGCCCATCGTGGGCGCCGAGACCTTCTTACGTGGAGCGCGCCCGCAACCGCG 1860  
2018 AGACCAAGATCGCAAGGCGCGGTACTGTGACCGACCGGGCGCGGCGCAAGATCGTGAGCC 2077  
Db AGACCAAGATCGGCAAGGCGCGGTACTGTGACCGACCGGGCGCGGCGCAAGGCTGTGAGCA 1920  
2078 TGACCGGAGACCAACCAAGAGACCGAGCTGCAAGGCGCATTCAGCTGGCGCTGCGAGACA 2137  
Db TCGCCGACACCAACCAAGAGACCGAGCTGCAAGGCGCATTCACCTGGCGCTGCGAGACA 1980  
2138 GCGGACGCGAGGTGAACATCGTGACCGACGACGAGTACGCCCTGGGGCATCATCCAGGCC 2197

1981 GCGGCTTGGAGGTGAACATCGTGACCCGACAGCCAGTAGTACGCCCTGGGCGATCATCCAGGCC 2040  
Qy AGCCCGCAAGAGAGCGAGAGGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAAG 2257  
Db AGCCCGCAAGAGAGCGAGAGGAGCTGGTGAAGCCAGATCATCGAGCAGCTGATCAAGAAAG 2100  
2258 AGAAGGTGTACTAGCTGGGTGCCCGCCCAAAAGGGGATCGCGGGCAACGAGCAGATCG 2317  
Db AGAAGGTGTACTAGCTGGGTGCCCGCCCAAAAGGGGATCGCGGGCAACGAGCAGGTGG 2160  
2318 ACAGCTGTGTAGCAAGGCGATCGCAAGTGTCTTCCAGAGGATCATCGATCGCGGCA 2377  
Db ACAGCTGTGTAGCAAGGCGATCGCAAGTGTCTTCCAGAGGATCATCGATCGCGGCA 2220  
2378 TCCTGATCTACCAAGTATCATGACGACCTGTACGTGGGCGAGCGGGCGCTTAGGATCGATT 2437  
Db TCCTGATCTACCAAGTATCATGACGACCTGTACGTGGGCGAGCGGGCGCTTAGGATCGATT 2280  
2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2469  
Db AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 7  
AX427930 9166 bp DNA linear PAT 20-JUN-2002  
LOCUS Sequence 168 from Patent WO0232943.  
DEFINITION AX427930  
ACCESSION AX427930  
VERSION AX427930.1 GI:21538017  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Huang, Y. and Nabel, G.J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 168 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES Location/Qualifiers  
source 1..9166  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 83.1%; Score 2051.2; DB 6; Length 9166;  
Best Local Similarity 92.6%; Pred. No. 8.2e-206;  
Matches 2192; Conservative 0; Mismatches 163; Indels 13; Gaps 3;

Qy 14 TGCGGAGGCGCATGAGCCAGCCAGCCAGCCAGCTCTGTATGTCAGCGGCAACTTCA 73  
Db TGCGGAGGCGCATGAGCCAGGTTGAACCAACCAACCATCATGATGACGCGCAACTTCA 3020  
Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGCAAGAGGGCGCACATCCCGCA 133  
Db AGGCCCCCAAGCGCACCATCAAGTGTCTTCAACTCGCGCAAGAGGGCGCACCTTGGCCCGCA 3080  
Qy 134 ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAAGTGGCGCAAGAGGGCGCCACAGATGA 193  
Db ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAAGTGGCGCAAGAGGGCGCCACAGGTGA 3140  
Qy 194 AGGACTGCAACGAGCGCCAGCCAACTTCTTCCGAGGAGCTTGGCTTCCCCCAGGGCA 253  
Db AGGACTGCAACGAGCGCCAGCGCTAA-TTTTTTAGGGAAGATCTTGGCTTCCCAAGGGA 3199  
Qy 254 AGGCCCCCGAGTTTCCCGAGCGAGCAACCGCGCCAAACAGCCCCCAGCGCGCTGC 313  
Db AGGCCCCCGAGTTTCTTCCAGAGCAGACGAGCCACAGCCACCAAGAGAGCTTC 3259  
Qy 314 AGGTGCGCGG-----CGACAAACCCCGCAGCGAGGCGCGCGCGGCGCA--- 364

[illegible]

4340	Db	ACAAGTGGAGCCGTGACGCCCATCGTGTGCTCCCGAGAAAGGACACGCTGGAACGCTGAACGACA	4399
1442	Qy	TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCGCCACGACAGATCTACCCCGGCATCAAGGTGC	1501
4400	Db	TCCAGAAGCTGGTGGGCAAGCTGAACTGGGCGACGACAGATCTACGCGCGGCATCAAGGTGC	4459
1502	Qy	GCAGCTGTGCAAAGCTGTCTGGCGGGCGCCAAAGCCCTGACCGACATCTGTGCCCTGACCG	1561
4460	Db	GCCAGCTGTGCAAAGCTGTCTGCGCGGCACCAAGAGCCCTGACCGAGGTGGTGTCCCTGACCG	4519
1562	Qy	AGGAGCGCAGCTGGAGCTGGCGCGAGAAACCGCCAGATCTCTGCGCGAGCCCGTGCACGGCG	1621
4520	Db	AGAGGCGGAGCTGGAGCTGGCGCGAGAACCGCCAGATCTCTGAAGGAGCCCGTGCACGGCG	4579
1622	Qy	TGTACTACGACCCACGACGACCTGTGTGCCCGAGATCCAGAAAGAGGGCCACGACCAAGT	1681
4580	Db	TGTACTACGACCCACGACGACCTGTATCGCCGAGATCCAGAAAGAGGGCCAGGGCCAGT	4639
1682	Qy	GGACCTACCAAGTCTTACGAGAGCCCTTCAGAAACCTGAAGACCGGCAAGTACGCCAAGA	1741
4640	Db	GGACCTACCAAGTCTTACGAGAGCCCTTCAAGAAACCTGAAGACCGGCAAGTACGCCCGCA	4699
1742	Qy	TGCGCACCGCCACACCAACGACGTGACGACCTGACCGAGGCGCTGCAGAAAGTACGCCA	1801
4700	Db	TGAAGGGCGGCCACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAAAGTACGCCA	4759
1802	Qy	TGAGAGCAGCTGTGATCTTGGGGCAAGACCCCAAGTTTCGCTTCGCCATCCAGAAAGGAGA	1861
4760	Db	CCGAGAGCATGTGATCTTGGGGCAGAACCCCAAGTTCAAGCTGCCCATCCAGAAAGGAGA	4819
1862	Qy	CTTGGGAGACCTGGTGGACCGATCTTGGCAGGCCACCTTGGATCTCCCGAGTGGGAGTTTCG	1921
4820	Db	CCTGGGAGGCTGGTGGACCGAGTACTTGGCAGGCCACCTTGGATCTCCCGAGTGGGAGTTTCG	4879
1922	Qy	TGAAACACCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCG	1981
4880	Db	TGAAACACCCCCCTGGTGAAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCGGCG	4939
1982	Qy	CCGAGACCTTCTACGTGGACGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGGCT	2041
4940	Db	CCGAGACCTTCTACGTGGACGGCGCGCCCAACCGCGAGACCAAGCTGGGCAAGGCCCGCT	4999
2042	Qy	ACGTGACCGACCGGGGC CGGACAGAGATCGTAGCTTGACCGAGACACCAACACAGAGA	2101
5000	Db	ACGTGACCGACCGCGCGCGCAGAAAGTGTGTGCCCTGACCGACACACCAACACAGAGA	5059
2102	Qy	CCGAGCTGAGGCCATTCAGCTGGCCCTGCAGACACGCGCAGGAGGTGAACATCGTGA	2161
5060	Db	CCGAGCTGAGGCCATTCACCTGGCCCTGCAGACACGCGCTGGAGGTGAACATCGTGA	5119
2162	Qy	CCGACGACGAGTACGCCCTGGGCATCATCCAGGCCCGACGCCCGACAGAGCGAGCGAGC	2221
5120	Db	CCGACGACGAGTACGCCCTGGGCATCATCCAGGCCCGACGCCCGACAGAGCGAGCGAGC	5179
2222	Qy	TGTTGAACCAAGATCATCGACAGCTGATCAAGAAAGAGAGGTGTACCTGAGCTGGGTGC	2281
5180	Db	TGTTGAGCCAGATCATCGACAGCTGATCAAGAAAGAGAGGTGTACCTGGCTGGGTGC	5239
2282	Qy	CCGCCCAAGGGCATCGGCGGCAACGAGCAGATTCGACAGCTGGTGCAGCAAGGGCATCC	2341
5240	Db	CCGCCCAAGGGCATCGGCGGCAACGAGCAGGTGGACGGCTTGGTAGCGCCGGCATCC	5299
2342	Qy	GCAAGTGTCTTCCTGGACGGCATCGA	2369
5300	Db	GCAAGTGTCTTCCTGGACGGCATCGA	5327
RESULT 8			
AX427936	LOCUS	AX427936	9788 bp DNA linear PAT 20-JUN-2002
DEFINITION	Sequence 174 from Patent WO0232943.		
ACCESSION	AX427936		
VERSION	AX427936.1 GI:21538023		

RESULT 8  
AX427936  
LOCUS  
DEFINITION  
ACCESSION  
VERSION



QY	1862	CTGGGAGACCTTGGTGGACCGA	CTACTGGCAGGCCACCTCGATCCCGAGTGGGAGTTGG	1921
Db	4815	CTTGGGAGGCTTGGTGGACCGAGT	ACTGGCAGGCCACCTCGATCCCGAGTGGGAGTTGG	4874
QY	1922	TGAACACCCGCCCTTGGTGAAGCT	GTGTGTAACAGCTGGAGAGGAGCCCATCATCGGGG	1981
Db	4875	TGAACACCCGCCCTTGGTGAAGCT	GTGTGTAACAGCTGGAGAGGAGCCCATCATCGGGG	4934
QY	1982	CCGAGACCTTCTACGTGGAGCGG	CGCCGACCAACCGCGAGACCAAGATCGGGCAGGCCGCT	2041
Db	4935	CCGAGACCTTCTACGTGGAGCGG	CGCCGACCAACCGCGAGACCAAGCTGGGCAAGGCCGCT	4994
QY	2042	ACGTGACCGACCGGGCCGCGACA	GATCGTGAAGCTGACCCGAGACCAACCAAGAGA	2101
Db	4995	ACGTGACCGACCGGGCCGCGACA	GATCGTGAAGCTGACCCGAGACCAACCAAGAGA	5054
QY	2102	CCGAGCTGCAAGGCCATCCAGCT	GGCCCTGCAGGACAGCGGACGCGAGGTGAACATCGTGA	2161
Db	5055	CCGAGCTGCAAGGCCATCCAGCT	GGCCCTGCAGGACAGCGGACGCGAGGTGAACATCGTGA	5114
QY	2162	CCGAGCCGAGTACGCCCTTGGG	CATCATCCAGGCCCGAGCCGACCAAGAGGAGGAGCGAGC	2221
Db	5115	CCGAGCCGAGTACGCCCTTGGG	CATCATCCAGGCCCGAGCCGACCAAGAGGAGGAGCGAGC	5174
QY	2222	TGCTGAACACAGATCATCGAG	CAGCTGATCAAGAGGAGGAGGTGTACCTGAGCTGGGTGC	2281
Db	5175	TGCTGAACACAGATCATCGAG	CAGCTGATCAAGAGGAGGAGGTGTACCTGAGCTGGGTGC	5234
QY	2282	CCGCCCAAGGGCATCGGCGG	CAACGAGCAGATCGCAACAGCTGGTGAGCAAGGGCATCC	2341
Db	5235	CCGCCCAAGGGCATCGGCGG	CAACGAGCAGATCGCAACAGCTGGTGAGCAAGGGCATCC	5294
QY	2342	GCAAGTGCTTCTTCTGGAGCG	GATCGATGGCGGCATCGTGATCTACCACTA	2393
Db	5295	GCAAGTGCTTCTTCTGGAGCG	GATCGATGGCGGCATCGTGATCTACCACTA	5346
RESULT 9				
AX427931				
LOCUS				
DEFINITION				
AX427931				
ACCESSION				
VERSION				
AX427931.1				
GI:21538018				
KEYWORDS				
SOURCE				
ORGANISM				
Huang, Y. and Nabel, G.J.				
Modifications of hiv env, gag, and pol enhance immunogenicity for				
genetic immunization				
Patent: WO 0232943-A 169 25-APR-2002;				
GOVERNMENT OF THE UNITED STATES (US)				
Location/Qualifiers				
1. .9169				
/organism="synthetic construct"				
/mol_type="unassigned DNA"				
/db_xref="taxon:32630"				
/notes="plasmid pVR1012x/s containing HIV genes"				
ORIGIN				
Query Match				
Best Local Similarity				
Matches 2194; Conservative				
82.2%; Score 2028.8; DB 6; Length 9169;				
Pred. No. 1.8e-203; Mismatches 182; Indels 16; Gaps 4;				
QY	14	TGCGCGAGGCGCATGAGCGAGCG	CACCCAGCGCCCAACATCTCTGATGCGAGCGGCAACTTCA	73
Db	2967	TGCGCGAGGCGCATGAGCGAGCG	CACCCAGCGCCCAACATCTCTGATGCGAGCGGCAACTTCA	3026
QY	74	AGGGCCCCAAGCGGCATCATCA	AGTGTGTAACCTGGCGCAAGGAGGGGCACATCGCCGCA	133
Db	3027	AGGGCCCCAAGCGGCATCATCA	AGTGTGTAACCTGGCGCAAGGAGGGGCACATCGCCGCA	3083

QY	134	ACTGCCGCGCCCCCGCAAGAGG	GCTGCTGGAAGTGCAGGAGGCGCACCAAGATGA	193
Db	3084	ACTGCCGCGCCCCCGCAAGAGG	GCTGCTGGAAGTGCAGGAGGCGCACCAAGATGA	3143
QY	194	AGGACTGCAACGAGGCCAGGCG	AACCTTCTTCGCGAGGACCTTGGCTTCCCCGAGGGCA	253
Db	3144	AGGACTGCAACGAGGCCAGGCG	AACCTTCTTCGCGAGGACCTTGGCTTCCCCGAGGGCA	3202
QY	254	AGGCGCGCGAGTTCCCGAGG	AGAGAACCGCGCCCAAGCCGCGAGCGCTGC	313
Db	3203	AGGCGCGCGAGTTCCCGAGG	AGAGAACCGCGCCCAAGCCGCGAGCGCTGC	3262
QY	314	AGGTGCGCGG-----CGACA	CCCCCGGAGGCGCGCGCGCGAGCGCAGGGCA---	364
Db	3263	AGGTGCGCGGAGAGACACAACT	CCCTCTCAGAACGAGGAGCGGATAGACAGGAACCTG	3322
QY	365	---CCCTGAACCTTCCCGAGAT	CAACCTCTGTGGAGCGCCCTCTGTGTGAGCATCAAGGTGG	421
Db	3323	TATCTCTTGTAGCTTCCCTCA	GATCACTCTTTGGCAGCGACCCCTCGTCAATAAAGATAG	3382
QY	422	GCGGCCAGATCAAGGAGGCGC	CTGCTGGACACCGCGCGCGACACCGCTGCTGGAGGAGA	481
Db	3383	GCGGCCAGCTGAAGGAGGCGC	CTTCTAGACACCGCGCGCGACACCGCTGCTGGAGGAGA	3442
QY	482	TGAGCTGCGCGCGCAAGTGA	AGCCCAAGATGATCGCGCGCGCATCGCGCGGCTTCAACAGG	541
Db	3443	TGAACTGCGCGCGCGCTGGA	AGCCCAAGATGATCGCGCGCGCATCGCGCGGCTTCAACAGG	3502
QY	542	TGCGCCAGTACGACACAGATC	CTGTGATCGAGATCTTGGCGCAAGAGGCGCATCGGCAACGTGC	601
Db	3503	TGCGCCAGTACGACACAGATC	CTGTGATCGAGATCTTGGCGCGCACCAAGGCGCATCGGCAACGTGC	3562
QY	602	TGATCGGCGCGCGCGCGGTA	ACATCATCGCGCGCAACATGCTGACCCAGCTCGGCTGCA	661
Db	3563	TGATCGGCGCGCGCGCGGTA	ACATCATCGCGCGCAACATGCTGACCCAGCTCGGCTGCA	3622
QY	662	CCCTGAACCTTCCCGCATCAG	CCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGCATGG	721
Db	3623	CCCTGAACCTTCCCGCATCAG	CCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGCATGG	3682
QY	722	ACGCGCCCCAAGGTGAAGCA	GTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCA	781
Db	3683	ACGCGCCCCAAGGTGAAGCA	GTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGCGTGAGA	3742
QY	782	TCCTGCGAGGAGATGGAGA	AGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACA	841
Db	3743	TCCTGCGAGGAGATGGAGA	AGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACA	3802
QY	842	ACACCCCGCTTTCGCGCATCA	AGAAGGAGACAGCACCAAGTGGCGCAAGCTGGTGGACT	901
Db	3803	ACACCCCGCTTTCGCGCATCA	AGAAGGAGACAGCACCAAGTGGCGCAAGCTGGTGGACT	3862
QY	902	TCGCGAGCTGAACAAGCGGA	CCCGAGCTTCTGGAGGTGCGAGCTGGGGCATCCCCCACC	961
Db	3863	TCGCGAGCTGAACAAGCGGA	CCCGAGCTTCTGGAGGTGCGAGCTGGGGCATCCCCCACC	3922
QY	962	CGCGCGCGCTGAGAGAGAG	AGAGCGTGCCTGGAAGCTGGGCGAGCGCTACTTTCA	1021
Db	3923	CGCGCGCGCTGAGAGAGAG	AGAGCGTGCCTGGAAGCTGGGCGAGCGCTACTTTCA	3982
QY	1022	GCGTGCCTTGGACGAGGACT	TTCCGCAAGTACACCGCTTTCACCTCCCGAGCATCAACA	1081
Db	3983	GCGTGCCTTGGACGAGGACT	TTCCGCAAGTACACCGCTTTCACCTCCCGAGCATCAACA	4042
QY	1082	ACAGAGACCCCGCGCATCCG	GTACAGTACAACTGTGCGCCGAGGCGTGAAGGCGAGCC	1141
Db	4043	ACAGAGACCCCGCGCATCCG	GTACAGTACAACTGTGCGCCGAGGCGTGAAGGCGAGCC	4102
QY	1142	CGAGCATCTTCAGAGCAGCA	TGACCAAGATCTCTGGAGCGCTTCCCGCGCCCGCAACCCCG	1201
Db	4103	CGAGCATCTTCAGAGCAGCA	TGACCAAGATCTCTGGAGCGCTTCCCGCGCCCGCAACCCCG	4162
QY	1202	AGATCGTGATCTACCAAGT	ACATGAGACCGCTGTACGTGGGCGAGCGACCTGGGATCGGCC	1261



Db 3506 AGTGGGCGAGTACGACAGATCTGTATCGAGATCTCGCGCCACAAAGGCCATCGGCACCG 3565  
QY 599 TGTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCT 658  
Db 3566 TGTGTTGGGCGCCACCCCGTGAACATCATCGCGCGCAACCTGCTGACCCAGATCGGCT 3625  
QY 659 GCACCTGGAATCTTCCCATCAGCCCATCAGAGCCGTCGCGGTGAAGCTGAAGCCCGGCA 718  
Db 3626 GCACCTGGAATCTTCCCATCAGCCCATCAGAGCCGTCGCGGTGAAGCTGAAGCCCGGCA 3685  
QY 719 TGGACCGCCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAAAGATCAAGGCCCTGACCG 778  
Db 3686 TGGACCGCCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAAAGATCAAGGCCCTGCTGG 3745  
QY 779 CCATCTGCGAGGAGATGGAAGAGGAGGGAAGATCACCAGATCGGGCCCGGAGAACCCCT 838  
Db 3746 AGATCTGCAACCGAGATGGAGAGGAGGGAAGATCAGCAAGATCGGCCCGCGAGAACCCCT 3805  
QY 839 ACAACACCCCGCTGTTCCGCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTG 898  
Db 3806 ACAACACCCCGCTGTTCCGCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTG 3865  
QY 899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGCTGGGCATCCCC 958  
Db 3866 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGCTGGGCATCCCC 3925  
QY 959 ACCCGCGCGCTGAAGAGAAAGAGCGTGAACCGTGTGAGCTGGGCGAGCCCTACT 1018  
Db 3926 ACCCGCGCGCTGAAGAGAAAGAGCGTGAACCGTGTGAGCTGGGCGAGCCCTACT 3985  
QY 1019 TCAGCGTGGCCCTGGACGAGACTTCCGCAAGTACACCGCTTCAACCATCCCCAGATCA 1078  
Db 3986 TCAGCGTGGCCCTGGACGAGACTTCCGCAAGTACACCGCTTCAACCATCCCCAGATCA 4045  
QY 1079 ACAACGAGACCCCGGCATCCGTTACAGTACAAACGTGTGCCCCAGGGCTGGAAGGCA 1138  
Db 4046 ACAACGAGACCCCGGCATCCGTTACAGTACAAACGTGTGCCCCAGGGCTGGAAGGCA 4105  
QY 1139 GCCCGAGCATCTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACC 1198  
Db 4106 GCCCGGCATCTTCCAGTGCAGATGACCAAGATCTTGGAGCCCTTCCGCAAGCAGAAC 4165  
QY 1199 CCAGAGTCTGTATCTACCATGACGACCTGTACGTGGCGAGCACTCGAGATCG 1258  
Db 4166 CCAGATCTGTATCTACCATGACGACCTGTACGTGGCGAGCACTCGAGATCG 4225  
QY 1259 GCAGCAACCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCGCTGGGGTTCAACCA 1318  
Db 4226 GCCAGCACCGCAAGATCGAGAGCTGCGCCAGCACCTGTGCGCTGGGGTTCAACCA 4285  
QY 1319 CCCCCAGACAGACCAAGAGGAGGCCCTTCTGTGGATGGGCTAGAGCTGCACC 1378  
Db 4286 CCCCCAGACAGACCAAGAGGAGGCCCTTCTGTGGATGGGCTAGAGCTGCACC 4345  
QY 1379 CCAGCAAGTGAACCGTGCACCCCATCGAGTGCCTGCGAGAGGAGAGCTGGAACGTAACG 1438  
Db 4346 CCAGCAAGTGAACCGTGCACCCCATCGAGTGCCTGCGAGAGGAGAGCTGGAACGTAACG 4405  
QY 1439 ACATCCAGAGCTGGTGGGCAAGCTGAATCTGGGCCAGCCAGATCTACCCCGCATCAAGG 1498  
Db 4406 ACATCCAGAGCTGGTGGGCAAGCTGAATCTGGGCCAGCCAGATCTACCCCGCATCAAGG 4465  
QY 1499 TGGCCAGCTGTCAAGCTGCTGCGCGCGCGCAAGGCCCTGACCGACATGTCGCCCTGA 1558  
Db 4466 TGGCCAGCTGTCAAGCTGCTGCGCGCGCGCAAGGCCCTGACCGAGGTGTCGCCCTGA 4525  
QY 1559 CCAGAGGCGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACG 1618  
Db 4526 CCAGAGGCGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGAGAGAGCCCGTGCACG 4585  
QY 1619 GCGTGTACTACGACCCCGCAAGAGACTGTGTGGCCGAGATCTCAGAGAGCGGCGACGACC 1678  
Db 4586 GCGTGTACTACGACCCCGCAAGAGACTGTGTGGCCGAGATCTCAGAGAGCGGCGAGGGCC 4645

QY 1679 AGTGGACCTACAGATCTTACCGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCA 1738  
Db 4646 AGTGGACCTACCGAGATCTTACCGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCC 4705  
QY 1739 AGATGCGCACCGGCCACACACGAGCTGAAGCAGCTGACCGAGGCGGTGACAGAGATCG 1798  
Db 4706 GCATGAAGGGCGGCCACACACGAGCTGAAGCAGCTGACCGAGGCGGTGACAGAGATCG 4765  
QY 1799 CMTGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGG 1858  
Db 4766 CCACCGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCAAGCTGCCCATCCAGAGG 4825  
QY 1859 AGACCTGGGAGACCTCTGTGACCGACTTACTGGCAGCGCACTTGATCCCGAGTGGGAGT 1918  
Db 4826 AGACCTGGGAGCGCTGTGTGACCGAGTACTGGCAGCGCACTTGATCCCGAGTGGGAGT 4885  
QY 1919 TCGTGAACACCCCGCTGTGTGAACTGTGTGATCCAGCTGGGAGAGGCCCATCATCG 1978  
Db 4886 TCGTGAACACCCCGCTGTGTGAACTGTGTGATCCAGCTGGGAGAGGCCCATCATCG 4945  
QY 1979 GCGCCAGACCTTCTAGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038  
Db 4946 GCGCCAGACCTTCTAGTGGAGCGCGCGCCCAACCGCGAGACCAAGTGGGCAAGGCCG 5005  
QY 2039 GCTACGTGACCGACCGGCGCGCGAGAGATCTGTGAGCTTGACCGAGACCAACCCAGA 2098  
Db 5006 GCTACGTGACCGACCGGCGCGCGAGAGTGTGTGCTGACCGAGACCAACCCAGA 5065  
QY 2099 AGACCGAGCTGCAAGGCCATCCAGCTGGCCCTTGACGACAGCGCGAGGTGAACATCG 2158  
Db 5066 AGACCGAGCTGCAAGGCCATCCAGCTGGCCCTTGACGACAGCGCGCTGGAGGTGAACATCG 5125  
QY 2159 TGAACGACAGCTGAGTGGCTGGGATCATTCAGGCGCCAGCCCGACAGAGCGAGAGCG 2218  
Db 5126 TGAACGACAGCTGAGTGGCTGGGATCATTCAGGCGCCAGCGAGCTGGAGCGCTGGAGCG 5185  
QY 2219 AGCTGTGTGAACCAAGTATCATCGAGCAGTATCAAGAGGAGAGGTGTACTCTGAGCTGGG 2278  
Db 5186 AGCTGTGTGACAGATCATCGAGCAGTATCAAGAGGAGAGGTGTACTCTGGCTGGG 5245  
QY 2279 TGCCCGCCCAAGAGGATCGGCGCAACGAGCAGATCGCAAGCTGGTGAAGAGGCA 2338  
Db 5246 TGCCCGCCCAAGAGGATCGGCGCAACGAGCAGTGGAGCGCTGGTGGAGCGCGGCA 5305  
QY 2339 TCGCAAGGTGCTGTTCTGGAGCGCATCGATGGCGGATCTGTGATCTACAGTA 2393  
Db 5306 TCGCAAGGTGCTGTTCTGGAGCGCATCGCAAGGCGCCAGGAGGAGCACAGAA 5360

## RESULT 11

AX427927 LOCUS 12411 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 165 from Patent WO0232943.  
ACCESSION AX427927  
VERSION AX427927.1 GI:21538014  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Huang, Y. and Nabel, G. J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 165 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES  
source  
1. .12411  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN



Query Match		82.1%; Score 2027; DB 6; Length 12411;
Best Local Similarity		91.6%; Pred. No. 2.6e-203;
Matches 2194; Conservative		0; Mismatches 185; Indels 16; Gaps 4;
Qy	14	TGGCCGAGGCGCATGAGCCAGG---CCACCAGCGCCACATCTCTGATCGAGCGCAGCAACT 70
Db	2967	TGGCCGAGGCGCATGAGCGAGGTGACCAACAGCGCCACCATCATGATGAGCGGCGCAACT 3026
Qy	71	TCAAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGCCACATCGCCC 130
Db	3027	TCCGCAACCGAGCGCAAGATCGTGAAGTCTTCAACTGCGGCAAGAGGCGCCACACCGCCC 3086
Qy	131	GCAACTGCGGCGCCCGCGCAAGAGGCTCTGGAAGTGGCGCAAGAGGCGCCACACAGA 190
Db	3087	GCAACTGCGGCGCCCGCGCAAGAGGCTCTGGAAGTGGCGCAAGAGGCGCCACACAGA 3146
Qy	191	TGAAGGACTGACACGAGCGCCAGGCGCAACTTCTTCCGCGAGGACTGCGCTTCCCGCAGG 250
Db	3147	TGAAGGACTGACACGAGCGGACAGGCTAA-TTTTTTAGGAAGATCTGCGCTTCCCAAG 3205
Qy	251	GCAAGGCGCGGAGTCTCCCGAGCGAGCAGAACCGCGCAACAGCCCCACACAGCGCGAGC 310
Db	3206	GGAAGGCCAGGGAATTTCTTCAGAGCAGACAGCCACAGCCCCACACAGAGAGAGC 3265
Qy	311	TGCAAGTGGCGG-----CGAACCCCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCA 364
Db	3266	TTCAAGTTTGGGGAAGAGACAACAACCTCCCTCTCAGAAGCAGAGGCGCGCATAGACAAGGAA 3325
Qy	365	-----CCCTGAACCTTCCCGAGATCACTCTGTCGGCAGCGCGCCCTGGTGAAGATCAAGG 418
Db	3326	CTGTATCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAACAATAAGA 3385
Qy	419	TGGCGCGCCAGATCAAGAGGCGCCCTGCTGGACACCGCGCGCGCGAGCAACCGTGTGGAGG 478
Db	3386	TAGGGGCGCAGCTGAAGAGGCGCTTCTAGACACCGGCGCGAGCAACCGTGTGGAGG 3445
Qy	479	AGATGAGCCTGCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTTCATCA 538
Db	3446	AGATGAACCTGCCCGCGCGCTGGAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTTCATCA 3505
Qy	539	AGGTGCGCCAGTAGCACAGATCTGTGATCGAGATCTGGGCGCAAGAGCCCATCGGACCG 598
Db	3506	AGGTGCGCCAGTAGCACAGATCTGTGATCGAGATCTGGGCGCAAGAGCCCATCGGACCG 3565
Qy	599	TGCTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCT 658
Db	3566	TGCTGTTGGGCGCCACCCCGTGAACATCATCGGCGCGCAACCTGCTGACCCAGATCGGCT 3625
Qy	659	GCACTCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGCA 718
Db	3626	GCACTCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGGCA 3685
Qy	719	TGGACGCGCCCAAGTGAAGAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGACCG 778
Db	3686	TGGACGCGCCCAAGTGAAGAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTGTTGG 3745
Qy	779	CCATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCGCCCGAGAACCCCT 838
Db	3746	AGATCTGCAACGAGATGGAGAGGAGGCAAGATCAGCAAGATCGGCGCCCGAGAACCCCT 3805
Qy	839	ACAACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCAACCAAGTGGCGCAAGCTGGTG 898
Db	3806	ACAACACCCCGTGTTCGCCATCAAGAAAGAGGACAGCAACCAAGTGGCGCAAGCTGGTG 3865
Qy	899	ACTTCCGCGAGCTGAACAGGCGACCCAGACTTCTGGAGGTGACGTGGGATCCCCC 958
Db	3866	ACTTCCGCGAGCTGAACAGGCGACCCAGACTTCTGGAGGTGACGTGGGATCCCCC 3925
Qy	959	ACCCCGCGGCTTGAAGAAAGAGAGCGGTGACCGTGTGGAGCTGGGCGAGCCCTACT 1018
Db	3926	ACCCCGCGGCTTGAAGAAAGAGAGCGGTGACCGTGTGGAGCTGGGCGAGCCCTACT 3985

Qy	1019	TCAGCGTGCCTCGACGAGGACTTCCGCAAGTACACCGCCCTTACCACTCCCGACATCA 1078
Db	3986	TCAGCGTGCCTCGACGAGGACTTCCGCAAGTACACCGCCCTTACCACTCCCGACATCA 4045
Qy	1079	ACAAAGAGACCCCGGATCCGCTACAGTACAACTGCTGCTCCCGAGGCTTGAAGGCA 1138
Db	4046	ACAAAGAGACCCCGGATCCGCTACAGTACAACTGCTGCTCCCGAGGCTTGAAGGCA 4105
Qy	1139	GCCCCAGCATTTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACC 1198
Db	4106	GCCCCAGCATTTTCCAGTGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCAAGCAACC 4165
Qy	1199	CCGAGATCTGTATCTACAGTACATGACGACCTGTAGCTGGGCGAGCGACTGGAGATCG 1258
Db	4166	CCGACATCTGTATCTACAGTACATGACGACCTGTAGCTGGGCGAGCGACTGGAGATCG 4225
Qy	1259	GCCAGACCGCGCCCAAGATCGAGGAGCTGCGCAAGCAGCTCTGCTGGCTGGGGCTTTCACCA 1318
Db	4226	GCCAGACCGCGCCCAAGATCGAGGAGCTGCGCAAGCAGCTCTGCTGGCTGGGGCTTTCACCA 4285
Qy	1319	CCCCCGACAAGACCAAGAGGAGCCCTTCTCTGTGTGATGGGCTTACGAGTGTCAACC 1378
Db	4286	CCCCCGACAAGACCAAGAGGAGCCCTTCTCTGTGTGATGGGCTTACGAGTGTCAACC 4345
Qy	1379	CCGACAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGCTGGGACCGTGAACG 1438
Db	4346	CCGACAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGCTGGGACCGTGAACG 4405
Qy	1439	ACATCCAGAAGCTGCTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGCTCAAGG 1498
Db	4406	ACATCCAGAAGCTGCTGGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGCTCAAGG 4465
Qy	1499	TGCGCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGGCGCTGACCGACATCTGTGCCCTTGA 1558
Db	4466	TGCGCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGGCGCTGACCGAGCTGTGCCCTTGA 4525
Qy	1559	CCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGATCTCTGGGCGAGCCCGTGCAGG 1618
Db	4526	CCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGATCTCTGAAGAGGCGCGTGCAGG 4585
Qy	1619	GGGTGTACTACGACCCCGAGCAAGGACCTGGTGGCGCGAGATCCAGAGCAGGCGCCAGACC 1678
Db	4586	GGGTGTACTACGACCCCGAGCAAGGACCTGGTGGCGCGAGATCCAGAGCAGGCGCCAGACC 4645
Qy	1679	AGTGAACCTTACAGATCTACCGAGGAGCCCTTCAAGAACCTTCAAGAACCTGGAAGACCGGCCA 1738
Db	4646	AGTGAACCTTACAGATCTACCGAGGAGCCCTTCAAGAACCTTCAAGAACCTGGAAGACCGGCCA 4705
Qy	1739	AGATGCGCACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCG 1798
Db	4706	GATGAAGGCGCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCG 4765
Qy	1799	CCATGGAGAGCATCGTGTATCTGGGCGAAGACCCCAAGTTCAAGTTCAAGTTCCCATTCAGAAAG 1858
Db	4766	CCATGGAGAGCATCGTGTATCTGGGCGAAGACCCCAAGTTCAAGTTCAAGTTCCCATTCAGAAAG 4825
Qy	1859	AGACTGGGAGACCTGGTGGACCGACTACTTGGCAGGCGCCACCTGGATCCCGAGTGGGAGT 1918
Db	4826	AGACTGGGAGACCTGGTGGACCGACTACTTGGCAGGCGCCACCTGGATCCCGAGTGGGAGT 4885
Qy	1919	TCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACAGAGTGGAGAGAGAGAGAGAGAGAGAGAG 1978
Db	4886	TCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACAGAGTGGAGAGAGAGAGAGAGAGAGAGAG 4945
Qy	1979	GCGCGGAGACCTTCTAGCTGAGCGGCGCCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
Db	4946	GCGCGGAGACCTTCTAGCTGAGCGGCGCCGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5005
Qy	2039	GCTACGTGAACCGAGCGGCGCGGAGAGAGTCTGTGAGCTGAGCTGAGAGAGAGAGAGAGAGAG 2098
Db	5006	GCTACGTGAACCGAGCGGCGCGGAGAGAGTCTGTGAGCTGAGCTGAGAGAGAGAGAGAGAGAGAG 5065
Qy	2099	AGACCGAGCTGACGCGCATCCAGCTGCGCTTGCAGGAGACAGCGGAGCGAGAGAGAGAGAGAG 2158



```
Db 5066 AGACCGAGCTGCAGGCATCCACCTGGCCCTGCAGGACAGCGGCTGGAGGTGAACATCG 5125
Qy 2159 TGACCGACAGCCAGTACGCCCTGGGATCATCCAGGCCACGCCGACAGAGCGAGCG 2218
Db 5126 TGACCGACAGCCAGTACGCCCTGGGATCATCCAGGCCACGCCGACAGAGCGAGCG 5185
Qy 2219 AGCTGGTGAACCAAGATCATCGACGAGTATCAAGAGGAGAGGTGTACTCGAGCTGGG 2278
Db 5186 AGCTGGTGAACCAAGATCATCGACGAGTATCAAGAGGAGAGGTGTACTCGGCTGGG 5245
Qy 2279 TGCCCGCCCAAGAGGCATCGCGGCCCAACGAGCAGATCGCAAGCTGGTGAGCAAGGGCA 2338
Db 5246 TGCCCGCCCAAGAGGCATCGCGGCCCAACGAGCAGGTGGACGCCGCTGGTGAGCGCCGCA 5305
Qy 2339 TCCGCAAGGTGCTGTCCTGGACGGATCGATCGCGGCATCGTGATCTACCACTA 2393
Db 5306 TCCGCAAGGTGCTGTCCTGGACGGATCGACAAAGGCCAGGAGGAGCACAGAA 5360

RESULT 12
LOCUS AX427926
DEFINITION Sequence 164 from Patent WO0232943.
ACCESSION AX427926
VERSION AX427926.1 GI:21538013
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Huang Y. and Nabel G.J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for
Genetic immunization
JOURNAL Patent: WO 0232943-A 164 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1..9194
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 82.0%; Score 2025.4; DB 6; Length 9194;
Best Local Similarity 91.6%; Pred. No. 4.1e-203;
Matches 2193; Conservative 0; Mismatches 186; Indels 16; Gaps 4;

Qy 14 TGGCCGAGGCCATGAGCGAGG---CCACACAGCGCCAACTCCTGTATGCGAGCGCAGCAACT 70
Db 2967 TGGCCGAGGCCATGAGCGAGGTGACCAACAGCGCCACCATCATGTATGCGAGCGGCACT 3026
Qy 71 TCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCGCCC 130
Db 3027 TCCGCAACACAGCGCAAGATCGTGAAGTGTCTTCAACTGGCGCAAGGAGGGCCACACGCC 3086
Qy 131 GCAACTGCGCGCGCCCGCCGCAAGAGGGCTGTCTGGAAGTCCGCGAGGAGGGCCACAGA 190
Db 3087 GCAACTGCGCGCGCCCGCCGCAAGAGGGCTGTCTGGAAGTCCGCAAGGAGGGCCACAGA 3146
Qy 191 TGAAGGACTGCAACGAGCGCCAGCGCAACTTCTTCGCGAGGACCTTGGCCCTTCCCCCAGG 250
Db 3147 TGAAGGACTGCAACGAGCGCACAGGCTTAA-TTTTTTAGGGAAGATCTGGCCCTTCCCAAG 3205
Qy 251 GCAAGGCCCGCGAGTTTCCCGAGCGAGCAGAACCGCGCCAAACAGCCCAACAGCGCGAGC 310
Db 3206 GGAAGGCCAGGGAATTCTTTCAGAGCAGACACAGACAGCCACAGCCACCCACCAAGAGAGC 3265
Qy 311 TGCAGGTGCGCG-----CGACAAACCCCGAGCGAGCGCGCGCGCGAGCGCGAGGCA 364
Db 3266 TTCAGGTTTGGGAAGAGAACAACTCCCTCTCAGAAAGCAGGAGCGGATAGACAAAGGAA 3325
Qy 365 -----CCCTGAACTTCCCCCAGATCACCCGTGTGGCAGCGCCCCCTGGTGAGCATCAAG 418
```

```
Db 3326 CTGTATCTCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAGA 3385
Qy 419 TGGGCGCGCAGATCAAGAGAGGCCCTCTGTGACACACCGCGCGCGACACACCGTCTGGAGG 478
Db 3386 TAGGGGGGCGAGCTGAAGAGAGGCCCTTCTAGACACCGCGCGCGACACACCGTCTGGAGG 3445
Qy 479 AGATGAGCCTGCCCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGGGCGGCTTCATCA 538
Db 3446 AGATGAACCTGCCCGCGCGCTGGAGCCCAAGATGATCGGGGCATCGGGCGGCTTCATCA 3505
Qy 539 AGTGCGCCAGGTACGACCAAGATCCTGATCGAGATCTGCGGCACAAAGGCGCATCGGCACCG 598
Db 3506 AGTGCGCCAGGTACGACCAAGATCCTGATCGAGATCTGCGGCACAAAGGCGCATCGGCACCG 3565
Qy 599 TGTGATCGGCGCCACCCCGGTGAACATCATCGGCGCGCAACATGCTGAGCCACGCTGGGCT 658
Db 3566 TGTGTTGGTGGCGCCACCCCGGTGAACATCATCGGCGCGCAACCTGCTGACCCAGATCGGCT 3625
Qy 659 GCACCTGGAATTTCCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCA 718
Db 3626 GCACCTGGAATTTCCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCA 3685
Qy 719 TGGACCGGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAAAGATCAAGGCCCTTGACCG 778
Db 3686 TGGACCGGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAAAGATCAAGGCCCTTGTTGG 3745
Qy 779 CMATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCAACCAAGATCGGCGCCCGAGAACCCCT 838
Db 3746 AGATCTGCAACCGAGATGGAGAGGAGGCGCAAGATCAGCAAGATCGGCGCCCGAGAACCCCT 3805
Qy 839 ACAACACCCCGTGTTCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGGTGG 898
Db 3806 ACAACACCCCGTGTTCGCCATCAAGAAAGAGGACAGACCAAGTGGCGCAAGCTGGTGG 3865
Qy 899 ACTTCCCGAGGTGAACAAGCGCACCCAGCGACTTCTGGGAGGTGCGAGCTGGGCATCCCC 958
Db 3866 ACTTCCCGAGGTGAACAAGCGCACCCAGCGACTTCTGGGAGGTGCGAGCTGGGCATCCCC 3925
Qy 959 ACCCGCGCGCCTGAAGAGAGAGAGCGGTGACCGTGTGCGACGTGGGCGAGCGCTACT 1018
Db 3926 ACCCGCGCGCCTGAAGAGAGAGAGCGGTGACCGTGTGCGACGTGGGCGAGCGCTACT 3985
Qy 1019 TCAGCTGTCCCTGGACAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCCAGCATCA 1078
Db 3986 TCAGCTGTCCCTGGACAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCA 4045
Qy 1079 ACAACGAGACCCCGGCATCCGCTACCAGTACAAACGTGTGCCCCAGGGGTGGAAGGGCA 1138
Db 4046 ACAACGAGACCCCGGCATCCGCTACCAGTACAAACGTGTGCCCCAGGGGTGGAAGGGCA 4105
Qy 1139 GCCCGAGCATTTCCGAGAGCAGATGACCAAGATCTTGGAGCCCTTCGCGCCCGCAACC 1198
Db 4106 GCCCGCCCATTTCCAGTGCAGATGACCAAGATCTTGGAGCCCTTCCCGAAGCAGAAC 4165
Qy 1199 CCGAGATCTGTATCTACCACTATCGACGACCTGTACGTGGGCGAGCACTCGAGATCG 1258
Db 4166 CCGCATCTGTATCTACCACTATCGACCACTGTACCTGTACGTGGGCGAGCACTCGAGATCG 4225
Qy 1259 GCCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGTGCGCTGGGGCTTCAACA 1318
Db 4226 GCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACTGTGCGCTGGGGCTTCAACA 4285
Qy 1319 CCCCCGACAAAGACCAAGAGAGGCCCTTCTGTGTGATGGGTAGAGCTGAGCTGACC 1378
Db 4286 CCCCCGACAAAGACCAAGAGAGGCCCTTCTGTGTGATGGGTAGAGCTGAGCTGAGCTGACC 4345
Qy 1379 CCGCAAGTGGACCGGTGACGCCCATCGAGCTGCGCGAGAGGAGAGCTGACCGCTGAACG 1438
Db 4346 CCGCAAGTGGACCGGTGACGCCCATCGTGTGCGCGAGAGGAGCAGCTGGAACG 4405
Qy 1439 ACATCAGAAAGCTGTGGGCAAGCTGAACTGTGGGCGCAGCCAGATCTTACCCCGGATCAAG 1498
```



```
QY 758 AGAAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCGCTGCTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660
QY 818 AGATCGGCCCCGAGAACCCCTCAACAACACCCCGTGTTCGCCATCAAGAAGAAGACAGCA 877
Db 661 AGATCGGCCCCGAGAACCCCTCAACAACACCCCGTGTTCGCCATCAAGAAGAAGACAGCA 720
QY 878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937
Db 721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780
QY 938 AGGTGCAGCTGGGCATCCCGCACCCCGCGCGCTGGAAGAAGAAGAAGAGGTGACCGTGC 997
Db 781 AGGTGCAGCTGGGCATCCCGCACCCCGCGCGCTGGAAGAAGAAGAAGAGGTGACCGTGC 840
QY 998 TGGACGTGGCGCAGCGCTACTTTCAGCGTGCCTTCGACGAGGACTTTCGCAAGTACACCG 1057
Db 841 TGGACGTGGCGCAGCGCTACTTTCAGCGTGCCTTCGACGAGGACTTTCGCAAGTACACCG 900
QY 1058 CTTTACCATCCCGAGATCAACAACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGC 1117
Db 901 CTTTACCATCCCGAGATCAACAACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGC 960
QY 1118 TGGCCAGAGGCTGGAAGGGCAGCCCGACATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db 961 TGGCCAGAGGCTGGAAGGGCAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
QY 1178 AGCCCTTTCGCGCGCGCAACCCCGAGATCGTGATCTACCAAGTACATGGACGACCTGTACG 1237
Db 1021 AGCCCTTTCGCGCGCGCAACCCCGAGATCGTGATCTACCA-----GGCCCCCTGTACG 1074
QY 1238 TGGGACGACCTGTGAGATCGGCAGCACCGCGCGCAAGATCGAGGAGCTGCGCAAGCAC 1297
Db 1075 TGGGACGACCTGTGAGATCGGCAGCACCGCGCAAGATCGAGGAGCTGCGCGAGCAC 1134
QY 1298 TGCTGGCTGGGCTTCACCAACCCCGACAAGAGCACCAAGAAGAGGAGCCCGCTTCTGT 1357
Db 1135 TGCTGGCTGGGCTTCACCAACCCCGACAAGAGCACCAAGAAGAGGAGCCCGCTTCTGT 1194
QY 1358 GGATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGGTGACGCCCATCGAGCTGCCGAGA 1417
Db 1195 GGATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGGTGACGCCCATCGAGCTGCCGAGA 1254
QY 1418 AGAGAGCTGACCGTGAAACGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCC 1477
Db 1255 AGGACAGCTGGACCGTGAAACGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCC 1314
QY 1478 AGATCTACCCCGGCATCAAGGTGCGCAGCTGTGCAAGCTGTGCGGCGGCCAAGGCC 1537
Db 1315 AGATCTACCCCGGCATCAAGGTGAGCAGCTGTGCAAGCTGTGCGGCGGCCAAGGCC 1374
QY 1538 TGACCGACATCGTGCCCTGACCGAGGAGCCGAGCTGGAGCTGGCGGAGAACCGGAGA 1597
Db 1375 TGACCGAGGTGATCCCTTACCGAGGAGCCGAGCTGGAGCTGGCGGAGAACCGGAGA 1434
QY 1598 TCTGCGGAGCCGTGCACGGGTGTACTACGACCCACAGAGGACCTGGTGGCGGAGA 1657
Db 1435 TCTGAAGGAGCCGTGCACGAGGTGTACTACGACCCACAGAGGACCTGGTGGCGGAGA 1494
QY 1658 TCCAGAAGCAGGCGCACGACAGTGGACCTTACAGATCTACAGGAGCCCTTCAAGAAC 1717
Db 1495 TCCAGAAGCAGGCGCACGAGCTGGACCTTACAGATCTACAGGAGCCCTTCAAGAAC 1554
QY 1718 TGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACACCAACGAGCTGAAGCAGCTGA 1777
Db 1555 TGAAGACCGGCAAGTACGCGCGCATGCGCGCGGCCACACCAACGAGCTGAAGCAGCTGA 1614
QY 1778 CCGAGGCGGTGCAGAAGATCGCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT 1837
Db 1615 CCGAGGCGGTGCAGAAGTGGAGCCGAGAGCATCGTGATCTGGGGCAAGATCCCAAGT 1674
```

```
QY 1838 TCGCCTTGCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGAGACCGACTACTTGGCAGGCCA 1897
Db 1675 TCAAGCTGCCCATCCAGAAAGGAGACCTGGGAGCCTGGTGAGTACTTGGCAGGCCA 1734
QY 1898 CTTGGATTCGCCGAGTGGGAGTTCGTGAACAACCCCCCCCCCTGGTGAAGCTGTGTACGAGC 1957
Db 1735 CTTGGATTCGCCGAGTGGGAGTTCGTGAACAACCCCCCCCCCTGGTGAAGCTGTGTACGAGC 1794
QY 1958 TGGAGAAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCG 2017
Db 1795 TGGAGAAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCG 1854
QY 2018 AGACCAAGATTCGGCAAGCGCGCTACGTGACCGACCGGGCGCGGAGAGATCGTGAGCC 2077
Db 1855 AGACCAAGCTGGGCAAGCGCGCTACGTGACCGACCGGGCGCGGAGAGTGTGTGAGCA 1914
QY 2078 TGAACGAGACCAACAACGAGACCGAGCTGACGCGCATCCAGCTGGGCCCTGCGAGGACA 2137
Db 1915 TCGCCGACACCAACAACGAGACCGAGCTGACGCGCATCCACCTGGGCCCTGCGAGGACA 1974
QY 2138 GCGGACGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGGCATCATCCAGGCC 2197
Db 1975 GCGGCTTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGGCATCATCCAGGCC 2034
QY 2198 AGCCGACAAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAG 2257
Db 2035 AGCCGACAAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAG 2094
QY 2258 AGAAGGTGTACTGAGCTGGTCCCGCCCAACAAGGCGATCGGCGGCAACGAGCAGATCG 2317
Db 2095 AGAAGGTGTACTGAGCTGGGTCGCGCCCAACAAGGCGATCGGCGGCAACGAGCAGGTG 2154
QY 2318 ACAAGCTGTGTGACAAAGGCGATCCGCAAGTGTCTTCTTGGACGCGCATCGATGGCGGCA 2377
Db 2155 ACAAGCTGTGTGACGCGCGCATCCGCAAGTGTCTTCTTGAACGCGCATCGATGGCGGCA 2214
QY 2378 TCGTGATCTACCAAGTACAGCAGCTGTGATGGGCGAGCGCGGCCCTTAGGATCGATT 2437
Db 2215 TCGTGATCTACCAAGTACAGCAGCTGTGATGGGCGAGCGCGGCCCTTAGGATCGATT 2274
QY 2438 AAAAGCTTCCGGGGCTAGCACCGGTGAATTC 2469
Db 2275 AAAAGCTTCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 14
CQ870574
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .2306
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="Description of Artificial Sequence:
FS(-).protmod.Rtopt.YM"

ORIGIN

Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 6.1e-203;
```

Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3	
Qy	170 GC GCGCAAGGAGGGCCACAGATGAAGGAGTGCACCGAGCGCGCAAGCTTCTTCCGCG 229
Db	1 GCGGCGCGAAGGACACCAATGAAGATTGCATCTGAGAGACAGGCTAAATTTCTTCCGCG 60
Qy	230 AGGACCTGGCCCTTCCGCCAGGGCAAGGCCCGGGAGTTCCCGAGCGAGGACGAGCAACCGCGCCA 289
Db	61 AGGACCTGGCCCTTCTCGAGGGCAAGGCCCGCGAGTTTCAGCAGGCGAGCAGACCGCGCGCA 120
Qy	290 ACAGCCCCACAGCGCGGAGCTGCAGGTGCGCGCG-----ACAAACCCCGCAGCGCAGG 343
Db	121 ACAGCCCAACCGCGCGAGCTGCAGGTGCGAGTGTGAGTGTGGCGGCGGAGAAACAAGCCTGAGCGAGG 180
Qy	344 CCGCGCGGAGCGCCAGGGCACCCCTG-----AACTTCCCGCAGATCAACCTGTGGCAGC 397
Db	181 CCGCGCGGACCGCGCAGGGCACCGTGAGCTTCAACTTCCCGCAGATCAACCTGTGGCAGC 240
Qy	398 GCCCCCTGTTGAGCATCAAGGTGGCGCGGCAGATCAAGGAGGCCCTGCTGAGACACCGCGG 451
Db	241 GCCCCCTGTTGAGCATCAAGGATCGCGCGGCAGCTCAAGGAGGCGCTGCTCGACACCGCGG 300
Qy	458 CCGACGACACCGTCTGTGAGGAGATGAGCTGCCCGCAAGTGTGAAGCCCAAGATGATCG 517
Db	301 CCGACGACACCGTCTGTGAGGAGATGAACCTGTCCCGGCAAGTGTGAAGCCCAAGATGATCG 360
Qy	518 GCGGCATCGCGCGCTTCATCAAGGTGGCGCCAGTACGACCAAGATCCTTGATCGAGATCTTGG 577
Db	361 GCGGCATCGCGGGCTTCATCAAGGTGCGGCAGTACGACCAAGATCCCGTGGAGATCTTGG 420
Qy	578 GCAAGAAAGGCATCGGCACCGTGTGATTCGGCCCCACCCCCCGTGAAACATCATATCGGCGCA 637
Db	421 GCCAAGGGCATCGGCACCGTGTGTGTGGCGCCACCCCCGTGAAACATCATATCGGCGCA 480
Qy	638 ACATGCTGACCCAGCTGGGCTGCACCTGTAACCTTCCCGCATCAGCCCGCATCGAGACCGTGC 697
Db	481 ACCTGCTGACCCAGATCGGCTGCAACCTGTAACCTTCCCGCATCAGCCCCCATCGAGACGCTGC 540
Qy	698 CCGTGAAGCTGAAGCCCGGCATGCACGGCCCCCAAGGTGAAGCAGTGGCCCTGTACCGCAGG 757
Db	541 CCGTGAAGCTGAAGCCGGGATGCACGGCCCCCAAGGTCAGCAGTGGCCCTGTACCGCAGG 600
Qy	758 AGAAGATCAAGGCCCTGACGCCCATCTCGAGGAGATGGAGAAGGAGGGCAAGATCACCA 817
Db	601 AGAAGATCAAGGCCCTGTGTGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA 660
Qy	818 AGATCGGCGCGGAGAAACCCCTAGAACACCCCGTGTTCGCCCATCAAGAAGAGACAGCA 877
Db	661 AGATCGGCGCGGAGAAACCCCTAGAACACCCCGTGTTCGCCCATCAAGAAGAGACAGCA 720
Qy	878 CCAAGTGGCGCAAGCTGGTGACTTCGCGAGCTTGAACAAGCGCACCCAGGACTTCTTGGG 937
Db	721 CCAAGTGGCGCAAGCTGGTGACTTCGCGAGCTTGAACAAGCGCACCCAGGACTTCTTGGG 780
Qy	938 AGGTGACGCTGGGCATCCCCACCCCGCGCCTGAAGAAAGAGAGCGTGACCGTGC 997
Db	781 AGGTGACGCTGGGCATCCCCACCCCGCGCCTGAAGAAAGAGAGCGTGACCGTGC 840
Qy	998 TGGAGTGGGGGAGCGCCTACTTTAGCGTGCCCTTGGACGAGGACTTTCGGCAAGTACACCG 1057
Db	841 TGGACGTGGGGCGAGCGCCTACTTTACGCGTGGCCCTTGGACAAGGACTTTCGGCAAGTACACCG 900
Qy	1058 CTTTACCATCCCGCAGCATCAACAACGAGACCCCGGCATCCGCTACAGTACAACTGC 1117
Db	901 CTTTCAACATCCCGCAGCATCAACAACGAGACCCCGGCATCCGCTACAGTACAACTGC 960
Qy	1118 TGCCCCAGGGCTGGAAGGGCAGGCCCAAGCATCTTTCAGAGCAGCATGACCAAGATCCTGG 1177
Db	961 TGCCCCAGGGCTGGAAGGGCAGGCCCGCCATCTTCAGAGCAGCATGACCAAGATCCTGG 1020
Qy	1178 AGCCCTTTCGCGCCCGCAACCCCGAGATCTGTGATCTACAGTACATGACGACCTGTACG 1237
Db	1021 AGCCCTTTCGCAAGCAGAAACCCGACATCTGTGATCTACCA-----GGCCCCCTGTACG 1074

Qy	1238	TG	GGCAGCAGC	ACTTGG	AGATCGG	CAGACA	CGCGCC	CAAGATCG	AGGAGCTCG	CAAGCACC	1239		
Db	1075	TG	GCAGCGCA	CCCTGG	AGATCGG	CAGACA	CGCGCA	CCCAAGATCG	AGGAGCTCG	CGCAGCACC	1134		
Qy	1298	TG	CTGGCCTTGG	GGCTTCA	CCACC	CCCGACA	AGAGAC	CAAGAGGAGCC	CCCTTCTCTGT		1357		
Db	1135	TG	CTGGCCTTGG	GGCTTCA	CCACC	CCCGACA	AGAGAC	CAAGAGGAGCC	CCCTTCTCTGT		1194		
Qy	1358	GG	ATGGGCTTAC	GAGCTGC	ACCC	CGACA	AAAGTGG	ACCGGTG	CAGCCCATCG	AGCTGCCCGAGA	1417		
Db	1195	GG	ATGGGCTTAC	GAGCTGC	ACCC	CGACA	AAAGTGG	ACCGGTG	CAGCCCATCG	AGCTGCCCGAGA	1254		
Qy	1418	AG	GAGAGCTCG	AACGTA	CCGTA	ACGA	CATC	CAGAAAGCTTGT	TGGGCAAGCTG	GAACCTTGGGCGCAGCC	1477		
Db	1255	AG	GACAGCTG	GAACGTA	CCGTA	ACGA	CATC	CAGAAAGCTTGT	TGGGCAAGCTG	GAACCTTGGGCGCAGCC	1314		
Qy	1478	AG	ATCTACCC	CGGCATCA	AGGTG	CGGCAG	CTGTG	CAAGCTGTG	CAAGCTGTG	CGCGCGCCCAAGGCCCC	1537		
Db	1315	AG	ATCTACCC	CGGCATCA	AGGTG	CGGCAG	CTGTG	CAAGCTGTG	CAAGCTGTG	CGCGCGCCCAAGGCCCC	1374		
Qy	1538	TG	ACCGACAT	CTGTG	CCCCCT	TGAC	CGAGGAG	CCGAGCTT	GGAGCTTGG	CGGCGAGAACCCGCGAGA	1597		
Db	1375	TG	ACCGAGGT	GTATCCCC	CTTGAC	CGAGGAG	CCGAGCTT	GGAGCTTGG	CGGCGAGAACCCGCGAGA	1434			
Qy	1598	TC	TGCGCGAG	CCCGTGC	ACGGG	GTGTACTA	CGACCC	CAAGCCCAAG	AGGACCTTGT	TGGCGCGAGA	1657		
Db	1435	TC	CTGAAGGAG	CCCGTGC	ACGGG	GTGTACTA	CGACCC	CAAGCCCAAG	AGGACCTTGT	TGGCGCGAGA	1494		
Qy	1658	TC	CAGAAGCA	AGGGCCAC	GACCTG	GACCTTA	CAGATCTT	ACAGGAG	CCCTTCA	AGAACCC	1717		
Db	1495	TC	CAGAAGCA	AGGGCCAC	GAGGCT	AGTGGAC	CTTAC	CAGATCTT	ACAGGAG	CCCTTCA	AGAACCC	1554	
Qy	1718	TG	AGACCG	GCAGTAGT	ACGCC	ACAGAT	CGCAC	CGCCCA	CCCAAC	AGAGCTG	GAACGAGCTGA	1777	
Db	1555	TG	AGACCG	GCAGTAGT	ACGCC	ACAGAT	CGCAC	CGCCCA	CCCAAC	AGAGCTG	GAACGAGCTGA	1614	
Qy	1778	CC	GAGGCGGT	GCAGAA	AGATCG	CCATG	CGAGAG	CACTCGT	GTATCTGG	GGGCAAGACCC	CCCAAGT	1837	
Db	1615	CC	GAGGCGGT	GCAGAA	AGGTG	AGCAC	CGAGAG	CACTCGT	GTATCTGG	GGGCAAGATCCCC	CAAGT	1674	
Qy	1838	TC	CGCTGCC	CTCCATTC	CAAGAG	GAGAC	CTTGG	TGGACCG	CACTTGG	CAGGCGCA	1897		
Db	1675	TC	AACTGTC	CCATTC	CAAGAG	GAGAC	CTTGG	GAGGACCTTGG	TGGATG	GGAGTACTGG	CAGGCGCA	1734	
Qy	1898	CT	TGGATCCC	CGAGTGG	GAAGTTC	GTGTA	CAACCC	CCCCCTTGG	TGTAAGCTGT	TGTGTAC	CAGC	1957	
Db	1735	CT	TGGATCCC	CGAGTGG	GAAGTTC	GTGTA	CAACCC	CCCCCTTGG	TGTAAGCTGT	TGTGTAC	CAGC	1794	
Qy	1958	TG	GAGAGG	AGGCCATCAT	CTGG	CGCGAG	ACCTTCTAC	GTGAG	CGGCGCGCC	CGCCCAACCGCG	2017		
Db	1795	TG	GAGAGG	AGGCCATCAT	CTGG	CGCGAG	ACCTTCTAC	GTGAG	CGGCGCGCC	CGCCCAACCGCG	1854		
Qy	2018	AG	ACCAAGAT	TCGG	CAAGG	CGGCTA	CGTGA	CCGAC	CGGGGCGG	GCAGAGATCGT	GAGCC	2077	
Db	1855	AG	ACCAAGCT	TGGGCA	AGGCGG	CTAC	GTGTAC	CCGAC	CGGGGCGG	GCAGAGATCGT	GAGCA	1914	
Qy	2078	TG	ACCGAG	ACCAAC	CAAGAG	ACCGAG	CTGCAG	GCCATCC	ACAGCTTGG	CGCCCTTCG	CAGGACA	2137	
Db	1915	TC	CCGAGCA	CCACCA	CCAGAG	ACCGAG	CTGCAG	GCCATCC	ACAGCTTGG	CGCCCTTCG	CAGGACA	1974	
Qy	2138	CG	GCAGCAG	AGGTGA	ACATCGT	GAC	CGAC	GAGCTAG	CGCCCTTGG	GCGATCATCC	AGGCCCC	2197	
Db	1975	GC	GGCTTGG	AGGTGA	ACATCGT	GAC	CGAC	GAGCTAG	CGCCCTTGG	GCGATCATCC	AGGCCCC	2034	
Qy	2198	AG	CCCAAG	AGCAG	CGAGCT	TGGTGA	ACAGATCAT	TCGAG	CAGCTGT	ATCAAG	AGG	2257	
Db	2035	AG	CCCAAG	AGCAG	CGAGCT	TGGTGA	ACAGATCAT	TCGAG	CAGCTGT	ATCAAG	AGG	2094	
Qy	2258	AGA	AGTGT	TACTG	AGCTTGG	TGCC	CGCCA	CAAGGG	GCATCG	CGCGCA	AGCAG	ATCG	2317
Db	2095	AGA	AGTGT	TACTG	AGCTTGG	TGCC	CGCCA	CAAGGG	GCATCG	CGCGCA	AGCAG	ATCG	2154

QY 2318 ACAAGCTGGTACGACAGGCGATCCGCAAGGTGCTGTTCTCTGGACGGCATCGATGGCGGCA 2377  
 Db 2155 ACAAGCTGGTACGACAGGCGGATCCGCAAGGTGCTGTTCTCTGGACGGCATCGATGGCGGCA 2214  
 QY 2378 TCGTGATCTACCAAGTATACGACGACCTGTACGTGGGACGGCGGCGCTTAGGATCGATT 2437  
 Db 2215 TCGTGATCTACCAAGTATACGACGACCTGTACGTGGGACGGCGGCGCTTAGGATCGATT 2274  
 QY 2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2469  
 Db 2275 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2306

RESULT 15  
 AR373387  
 LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003  
 DEFINITION Sequence 82 from patent US 6602705.  
 ACCESSION AR373387  
 VERSION AR373387.1 GI:40075490  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 2306)  
 AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
 TITLE Expression of HIV polypeptides and production of virus-like particles  
 JOURNAL Patent: US 6602705-A 82 05-AUG-2003;  
 FEATURES  
 source Location/Qualifiers  
 1..2306  
 /organism="unknown"  
 /mol\_type="genomic DNA"

ORIGIN  
 Query Match 82.0%; Score 2025.2; DB 6; Length 2306;  
 Best Local Similarity 93.3%; Pred. No. 6.1e-203;  
 Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
 QY 170 GCGCAGAGGAGCCACCAAGTGAAGACGTGCAACGAGCCGAGGCCCACTTCTCCGCG 229  
 Db 1 GCGGCGCGAAGGACCAAAATGAAGATTGCACTGAGACAGGCTAATTCTTCCGCG 60  
 QY 230 AGGACCTTGGCTTCCCGCAGGCAAGCCCGCGAGTTTCCCGAGCGAGCAAGCCGCGCCA 289  
 Db 61 AGGACCTTGGCTTCCCGCAGGCAAGCCCGCGAGTTTCCCGAGCGAGCAAGCCGCGCCA 120  
 QY 290 ACAGCCCCACAGCGCGAGCTGAGGTGCGCGCG-----ACAACCCCGCGAGG 343  
 Db 121 ACAGCCCCACCGCGCGAGCTGAGGTGCGCGCGCGCGAGAAACAACAGCTGAGCGAGG 180  
 QY 344 CCGGCGCGAGCGCGAGCG 397  
 Db 181 CCGGCGCGAGCGCGCGAGCG 240  
 QY 398 GCCCCTGTGTAGCATCAAGGTGGCGCGCGCGAGTCAAGGAGGCGCTGCTGCAACCGCGCG 457  
 Db 241 GCCCCTGTGTAGCATCAAGGTGGCGCGCGCGAGTCAAGGAGGCGCTGCTGCAACCGCGCG 300  
 QY 458 CCGACGACACCGT 517  
 Db 301 CCGACGACACCGT 360  
 QY 518 CCGGCGCGAGCG 577  
 Db 361 CCGGCGCGAGCG 420  
 QY 578 GCAAGAGGCGCATCGGCGACCGT 637  
 Db 421 GCAAGAGGCGCATCGGCGACCGT 480  
 QY 638 ACATGTGACCCAGCTGGGTGCAACCTGAACTTCCCGCATCAGCCCCCATCAGACCGGTGC 697  
 Db 481 ACCTGTGTACCCAGATCGGCTGCACCTTGAATTTCCCGCATCAGCCCCCATCAGACCGGTGC 540

QY 698 CCGTGAAGCTGAAGCCCGCGCATGGAAGCGGCCCAAGGTGAAGCAGTGGCCCTTACCGAGG 757  
 Db 541 CCGTGAAGCTGAAGCCCGGGATGGAAGCGGCCCAAGGTCAAGCAGTGGCCCTTACCGAGG 600  
 QY 758 AGAAGATCAAGGCGCTTGAACCGCATCTGCGAGGAGATGGAAGAGGAGGCGCAAGATCACCA 817  
 Db 601 AGAAGATCAAGGCGCTTGGTGGAGATCTGCACCGAGATGGAGAAGAGGCGCAAGATCAGCA 660  
 QY 818 AGATCGGCCCGGAGAACCCCTTCAACACACCCCGGTGTTCCGCATCAAGAGAAGAGGAGCA 877  
 Db 661 AGATCGGCCCGGAGAACCCCTTCAACACACCCCGGTGTTCCGCATCAAGAGAAGAGGAGCA 720  
 QY 878 CCAAGTGGCGCAAGCTGGTGGACTTCCGGGAGCTGAAACAAGCGCACCCAGGACTTCTTGGG 937  
 Db 721 CCAAGTGGCGCAAGCTGGTGGACTTCCGGGAGCTGAAACAAGCGCACCCAGGACTTCTTGGG 780  
 QY 938 AGGTGAGCTGGGCGATCCCCCAGCCCGCGGCTGAAAGAGAGAGAGGAGGAGCGTGC 997  
 Db 781 AGGTGAGCTGGGCGATCCCCCAGCCCGCGGCTGAAAGAGAGAGGAGGAGCGTGC 840  
 QY 998 TGGACGTGGCGAGCGCTACTTTCAGCGTGGCCCTGGAACGAGGAGCTTCCGCAAGTACCG 1057  
 Db 841 TGGACGTGGCGAGCGCTACTTTCAGCGTGGCCCTGGAACGAGGAGCTTCCGCAAGTACCG 900  
 QY 1058 CTTTACCACATCCCGAGCATCAACAAGAGAGACCCCGGCTCCGCTACCAAGTACAGTGC 1117  
 Db 901 CTTTACCACATCCCGAGCATCAACAAGAGAGACCCCGGCTCCGCTACCAAGTACAGTGC 960  
 QY 1118 TGCCCCAGGCTGGAAGGAGCGCCCGAGCATTTTCAGAGCAGCATGACCAAGATCTGTGG 1177  
 Db 961 TGCCCCAGGCTGGAAGGAGCGCCCGGCGCATTTTCAGAGCAGCATGACCAAGATCTGTGG 1020  
 QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATCGTGTATCTTACAGTATACATGGAGCAGCTGACG 1237  
 Db 1021 AGCCCTTCCGCGAAGCAGAACCCCGCATCGTGTATCTACCA-----GGCCCCCTGTGACG 1074  
 QY 1238 TGGCAGCAGCACTTGGAGATCGGCCAGCACCGCGCCAAAGATCCGAGGAGCTGCGCAAGCACC 1297  
 Db 1075 TGGCAGCAGCACTTGGAGATCGGCCAGCACCGCACCAAGATCCGAGGAGCTGCGCCAGCACC 1134  
 QY 1298 TGTGCGCTGGGCTTTCACACCCCGCAAGAAGCAACAGAGAGGAGCGCCCTTCTGT 1357  
 Db 1135 TGTGCGCTGGGCTTTCACACCCCGCAAGAAGCAACAGAGGAGCGCCCTTCTGT 1194  
 QY 1358 GGTATGGCTACGAGCTGCAACCGCAAGTGGACCGTGGAGCCCATCGAGCTGCCCGGAGA 1417  
 Db 1195 GGTATGGCTACGAGCTGCAACCGCGCAAGTGGACCGTGGAGCCCATCTGCTGCCCGAGA 1254  
 QY 1418 AGGAGAGCTGGACCGGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAAGCTGGGCGCAGCC 1477  
 Db 1255 AGGAGAGCTGGACCGGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAAGCTGGGCGCAGCC 1314  
 QY 1478 AGATTCACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGGCGGCGCAAGGCCC 1537  
 Db 1315 AGATTCACCCCGCATCAAGGTGAAGCTGTGCAAGCTGTGCGGCGGCGCAAGGCCC 1374  
 QY 1538 TGAACGACATCGTGGCCCTGACCGAGGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGA 1597  
 Db 1375 TGAACGAGGTGATTCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGA 1434  
 QY 1598 TCTGTGCGGAGCGCGGTGCAACCGGTGTACTTCAACGCCAGCAAGGAGCCTGGTGGCGCGAGA 1657  
 Db 1435 TCTGTGAAGAGCGCGGTGCAACCGGTGTACTTCAACGCCAGCAAGGAGCCTGGTGGCGCGAGA 1494  
 QY 1658 TCAGAAGCAGGCGCCACGACCAAGTGAACCTTACAGATCTTACAGGAGCGCTTCAAGAAC 1717  
 Db 1495 TCAGAAGCAGGCGCCACGAGCGAGTGAACCTTACAGATCTTACAGGAGCGCTTCAAGAAC 1554  
 QY 1718 TGAAGACCGGCGAGTACGCGCAGATGGGACCGCGCCACCAACGACGCTGAACGAGCTGA 1777  
 Db 1555 TGAAGACCGGCGAGTACGCGCGCGCATGCGCGGCGCCCAACGACGCTGAACGAGCTGA 1614

Qy	1778	CCGAGGCGCTGCAGAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGT	1837
Db	1615		1674
Qy	1838	TCCGCTGCCCATCCAGAAAGAGACTCGGAGACCTGGTGACCGACTACTGGCAGGCCA	1897
Db	1675		1734
Qy	1898	CCTGGATCCCAGAGTGGAGTTCTGTGAACACCCCGCCCTCGTGAAGCTGTGTACCAAGC	1957
Db	1735		1794
Qy	1958	TGGAGAAGAGCCCATCATCGCGGCCGAGACCTTCTAGTGGACGGCGCCGCCAACCGCG	2017
Db	1795		1854
Qy	2018	AGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAAATCGTGAGCC	2077
Db	1855		1914
Qy	2078	TGACCGAGACCAACCAACAGAAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA	2137
Db	1915		1974
Qy	2138	GCGGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCC	2197
Db	1975		2034
Qy	2198	AGCCCGACAGCGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGG	2257
Db	2035		2094
Qy	2258	AGAAAGGTGTACTTGAGCTGGGTGCCCGCCCAAGGGCATCGGGCGCAACGAGCAGATCG	2317
Db	2095		2154
Qy	2318	ACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTTGAGCGGCATCGATGGCGGCA	2377
Db	2155		2214
Qy	2378	TCGTGATCTACCACTACATGACGACCTGTACGTGGCGAGCGCGCCCTAGGATCGATT	2437
Db	2215		2274
Qy	2438	AAAAGCTTCCCGGGCTAGCACCGGTGAATTC	2469
Db	2275		2306

Search completed: June 1, 2005, 20:20:25  
Job time : 10578.3 secs

***This Page Blank (uspto)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 10:29:57 ; Search time 1414.77 Seconds  
(without alignments)  
10728.225 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469

Sequence: 1 gtcagccaccatggccga.....gggctagaccgtgaattc 2469

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5706582 seqs, 3073711274 residues

Total number of hits satisfying chosen parameters: 11413164

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq:\*
- 19: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 20: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq:\*
- 21: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 22: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	2469	100.0	2469	10	US-09-899-575-30
2	2457	99.5	2457	15	US-10-190-435-45
3	2457	99.5	2457	16	US-10-190-305A-39
4	2442.2	98.9	2463	10	US-09-899-575-31
5	2434.6	98.6	2457	15	US-10-190-435-44
6	2434.6	98.6	2457	16	US-10-190-305A-38
7	2415.4	97.8	2457	10	US-09-899-575-32
8	2401.8	97.3	2445	15	US-10-190-435-43
9	2401.8	97.3	2445	16	US-10-190-305A-37
10	2394.8	97.0	3930	15	US-10-190-435-9
11	2393.2	96.9	3930	15	US-10-190-435-10
12	2393.2	96.9	3930	15	US-10-190-435-11
13	2393.2	96.9	3930	15	US-10-190-435-12
14	2393.2	96.9	3930	15	US-10-190-435-13
15	2393.2	96.9	3930	15	US-10-190-435-14
16	2393.2	96.9	3930	15	US-10-190-435-15
17	2393.2	96.9	3930	15	US-10-190-435-16
18	2393.2	96.9	3930	15	US-10-190-435-17
19	2393.2	96.9	3930	15	US-10-190-435-18
20	2393.2	96.9	3930	15	US-10-190-435-19
21	2393.2	96.9	3930	15	US-10-190-435-20
22	2393.2	96.9	3930	15	US-10-190-435-21
23	2393.2	96.9	3930	15	US-10-190-435-22
24	2393.2	96.9	3930	15	US-10-190-435-23
25	2393.2	96.9	3930	15	US-10-190-435-24
26	2393.2	96.9	3930	15	US-10-190-435-25
27	2393.2	96.9	3930	15	US-10-190-435-26
28	2393.2	96.9	3930	15	US-10-190-435-27
29	2393.2	96.9	3930	15	US-10-190-435-28
30	2393.2	96.9	3930	15	US-10-190-435-29
31	2393.2	96.9	3930	15	US-10-190-435-30
32	2393.2	96.9	3930	15	US-10-190-435-31
33	2393.2	96.9	3930	15	US-10-190-435-32
34	2393.2	96.9	3930	15	US-10-190-435-33
35	2393.2	96.9	3930	15	US-10-190-435-34
36	2393.2	96.9	3930	15	US-10-190-435-35
37	2393.2	96.9	3930	15	US-10-190-435-36
38	2393.2	96.9	3930	15	US-10-190-435-37
39	2393.2	96.9	3930	15	US-10-190-435-38
40	2393.2	96.9	3930	15	US-10-190-435-39
41	2393.2	96.9	3930	15	US-10-190-435-40
42	2393.2	96.9	3930	15	US-10-190-435-41
43	2393.2	96.9	3930	15	US-10-190-435-42
44	2393.2	96.9	3930	15	US-10-190-435-43
45	2393.2	96.9	3930	15	US-10-190-435-44
46	2393.2	96.9	3930	15	US-10-190-435-45
47	2393.2	96.9	3930	15	US-10-190-435-46
48	2393.2	96.9	3930	15	US-10-190-435-47
49	2393.2	96.9	3930	15	US-10-190-435-48
50	2393.2	96.9	3930	15	US-10-190-435-49
51	2393.2	96.9	3930	15	US-10-190-435-50
52	2393.2	96.9	3930	15	US-10-190-435-51
53	2393.2	96.9	3930	15	US-10-190-435-52
54	2393.2	96.9	3930	15	US-10-190-435-53
55	2393.2	96.9	3930	15	US-10-190-435-54
56	2393.2	96.9	3930	15	US-10-190-435-55
57	2393.2	96.9	3930	15	US-10-190-435-56
58	2393.2	96.9	3930	15	US-10-190-435-57
59	2393.2	96.9	3930	15	US-10-190-435-58
60	2393.2	96.9	3930	15	US-10-190-435-59
61	2393.2	96.9	3930	15	US-10-190-435-60
62	2393.2	96.9	3930	15	US-10-190-435-61
63	2393.2	96.9	3930	15	US-10-190-435-62
64	2393.2	96.9	3930	15	US-10-190-435-63
65	2393.2	96.9	3930	15	US-10-190-435-64
66	2393.2	96.9	3930	15	US-10-190-435-65
67	2393.2	96.9	3930	15	US-10-190-435-66
68	2393.2	96.9	3930	15	US-10-190-435-67
69	2393.2	96.9	3930	15	US-10-190-435-68
70	2393.2	96.9	3930	15	US-10-190-435-69
71	2393.2	96.9	3930	15	US-10-190-435-70
72	2393.2	96.9	3930	15	US-10-190-435-71
73	2393.2	96.9	3930	15	US-10-190-435-72
74	2393.2	96.9	3930	15	US-10-190-435-73
75	2393.2	96.9	3930	15	US-10-190-435-74
76	2393.2	96.9	3930	15	US-10-190-435-75
77	2393.2	96.9	3930	15	US-10-190-435-76
78	2393.2	96.9	3930	15	US-10-190-435-77
79	2393.2	96.9	3930	15	US-10-190-435-78
80	2393.2	96.9	3930	15	US-10-190-435-79
81	2393.2	96.9	3930	15	US-10-190-435-80
82	2393.2	96.9	3930	15	US-10-190-435-81
83	2393.2	96.9	3930	15	US-10-190-435-82
84	2393.2	96.9	3930	15	US-10-190-435-83
85	2393.2	96.9	3930	15	US-10-190-435-84
86	2393.2	96.9	3930	15	US-10-190-435-85
87	2393.2	96.9	3930	15	US-10-190-435-86
88	2393.2	96.9	3930	15	US-10-190-435-87
89	2393.2	96.9	3930	15	US-10-190-435-88
90	2393.2	96.9	3930	15	US-10-190-435-89
91	2393.2	96.9	3930	15	US-10-190-435-90
92	2393.2	96.9	3930	15	US-10-190-435-91
93	2393.2	96.9	3930	15	US-10-190-435-92
94	2393.2	96.9	3930	15	US-10-190-435-93
95	2393.2	96.9	3930	15	US-10-190-435-94
96	2393.2	96.9	3930	15	US-10-190-435-95
97	2393.2	96.9	3930	15	US-10-190-435-96
98	2393.2	96.9	3930	15	US-10-190-435-97
99	2393.2	96.9	3930	15	US-10-190-435-98
100	2393.2	96.9	3930	15	US-10-190-435-99
101	2393.2	96.9	3930	15	US-10-190-435-100
102	2393.2	96.9	3930	15	US-10-190-435-101
103	2393.2	96.9	3930	15	US-10-190-435-102
104	2393.2	96.9	3930	15	US-10-190-435-103
105	2393.2	96.9	3930	15	US-10-190-435-104
106	2393.2	96.9	3930	15	US-10-190-435-105
107	2393.2	96.9	3930	15	US-10-190-435-106
108	2393.2	96.9	3930	15	US-10-190-435-107
109	2393.2	96.9	3930	15	US-10-190-435-108
110	2393.2	96.9	3930	15	US-10-190-435-109
111	2393.2	96.9	3930	15	US-10-190-435-110
112	2393.2	96.9	3930	15	US-10-190-435-111
113	2393.2	96.9	3930	15	US-10-190-435-112
114	2393.2	96.9	3930	15	US-10-190-435-113
115	2393.2	96.9	3930	15	US-10-190-435-114
116	2393.2	96.9	3930	15	US-10-190-435-115
117	2393.2	96.9	3930	15	US-10-190-435-116
118	2393.2	96.9	3930	15	US-10-190-435-117
119	2393.2	96.9	3930	15	US-10-190-435-118
120	2393.2	96.9	3930	15	US-10-190-435-119
121	2393.2	96.9	3930	15	US-10-190-435-120
122	2393.2	96.9	3930	15	US-10-190-435-121
123	2393.2	96.9	3930	15	US-10-190-435-122
124	2393.2	96.9	3930	15	US-10-190-435-123
125	2393.2	96.9	3930	15	US-10-190-435-124
126	2393.2	96.9	3930	15	US-10-190-435-125
127	2393.2	96.9	3930	15	US-10-190-435-126
128	2393.2	96.9	3930	15	US-10-190-435-127
129	2393.2	96.9	3930	15	US-10-190-435-128
130	2393.2	96.9	3930	15	US-10-190-435-129
131	2393.2	96.9	3930	15	US-10-190-435-130
132	2393.2	96.9	3930	15	US-10-190-435-131
133	2393.2	96.9	3930	15	US-10-190-435-132
134	2393.2	96.9	3930	15	US-10-190-435-133
135	2393.2	96.9	3930	15	US-10-190-435-134
136	2393.2	96.9	3930	15	US-10-190-435-135
137	2393.2	96.9	3930	15	US-10-190-435-136
138	2393.2	96.9	3930	15	US-10-190-435-137
139	2393.2	96.9	3930	15	US-10-190-435-138
140	2393.2	96.9	3930	15	US-10-190-435-139
141	2393.2	96.9	3930	15	US-10-190-435-140
142	2393.2	96.9	3930	15	US-10-190-435-141
143	2393.2	96.9	3930	15	US-10-190-435-142
144	2393.2	96.9	3930	15	US-10-190-435-143
145	2393.2	96.9	3930	15	US-10-190-435-144
146	2393.2	96.9	3930	15	US-10-190-435-145
147	2393.2	96.9	3930	15	US-10-190-435-146
148	2393.2	96.9	3930	15	US-10-190-435-147
149	2393.2	96.9	3930	15	US-10-190-435-148
150	2393.2	96.9	3930	15	US-10-190-435-149

QY	61	CCGAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC	120
Db	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC	120
QY	121	CACATGCGCCGCAACTGCGCGCCCCCGCCGCAAGAAAGGCTGTGGAAGTGCAGCAAGGAG	180
Db	121	CACATGCGCCGCAACTGCGCGCCCCCGCCGCAAGAAAGGCTGTGGAAGTGCAGCAAGGAG	180
QY	181	GGCCACAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACTGTGGCC	240
Db	181	GGCCACAGATGAAGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACTGTGGCC	240
QY	241	TTCCCCCAGGCAAGGCCCCCGAGTTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Db	241	TTCCCCCAGGCAAGGCCCCCGAGTTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
QY	301	AGCCGGAGCTGAGGTGCGCGCGCAACACCCCGCAGGAGCGCCGCGCGAGCGCCAG	360
Db	301	AGCCGGAGCTGAGGTGCGCGCGCAACACCCCGCAGGAGCGCCGCGCGAGCGCCAG	360
QY	361	GGCACCTGAACTTCCCCCAGATCACCTGTGTGCGAGGCCCCCTGTGTGAGCATCAAGGTG	420
Db	361	GGCACCTGAACTTCCCCCAGATCACCTGTGTGCGAGGCCCCCTGTGTGAGCATCAAGGTG	420
QY	421	GGCGGCCAGATCAAGGAGGCCCTGTGTGACACCGGCGCCGACGACACCGTGTGAGGAG	480
Db	421	GGCGGCCAGATCAAGGAGGCCCTGTGTGACACCGGCGCCGACGACACCGTGTGTGAGGAG	480
QY	481	ATGAGCCTTGC CGCGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGGCTTTCATCAAG	540
Db	481	ATGAGCCTTGC CGCGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGGCTTTCATCAAG	540
QY	541	GTGCGCCAGTACGACGATCTGTGATCGAGATCTGTGGGCAAGAGGCGCATCGGCAACG	600
Db	541	GTGCGCCAGTACGACGATCTGTGATCGAGATCTGTGGGCAAGAGGCGCATCGGCAACG	600
QY	601	CTGATCGGCCCCACCCCGTGAAATCATCATCGCGCGCAACATGTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCCCCACCCCGTGAAATCATCATCGCGCGCAACATGTGACCCAGCTGGGCTGC	660
QY	661	ACCTTGAATTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGTGAAGCCCGGCATG	720
Db	661	ACCTTGAATTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGTGAAGCCCGGCATG	720
QY	721	GACGCCCCAAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780
Db	721	GACGCCCCAAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780
QY	781	ATCTGCGAGGAGTGGAGAGGAGGGCAAGATCACCAAGATCGGCCCGCAGAAACCCCTAC	840
Db	781	ATCTGCGAGGAGTGGAGAGGAGGGCAAGATCACCAAGATCGGCCCGCAGAAACCCCTAC	840
QY	841	AACACCCCGCTGTTCGCATCAAGAAAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Db	841	AACACCCCGCTGTTCGCATCAAGAAAGGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900
QY	901	TTCCGGAGCTGAAACAGGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCGATCCCCAC	960
Db	901	TTCCGGAGCTGAAACAGGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCGATCCCCAC	960
QY	961	CCGCGCGCTGAAGAGAGAGAGCGTGACCGTGTGACCTGGGCGAGCGCTACTTC	1020
Db	961	CCGCGCGCTGAAGAGAGAGAGCGTGACCGTGTGACCTGGGCGAGCGCTACTTC	1020
QY	1021	AGGCTGCCCCCGCAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCCCAGCATCAAC	1080
Db	1021	AGGCTGCCCCCGCAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCCCAGCATCAAC	1080
QY	1081	AACGAGACCCCGGCATCCGCTACAGTACAAAGTGTGCCCCCAGGGCTGGAAGGGGAGC	1140
Db	1081	AACGAGACCCCGGCATCCGCTACAGTACAAAGTGTGCCCCCAGGGCTGGAAGGGGAGC	1140
QY	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCGCCCAACCCC	1200

Db	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCGCCCAACCCC	1200
QY	1201	GAGATCGTGTATCTTACAGTACATGGAGCAGCTGTAGCTGGGCAAGCAGCTGGAGATCGGC	1260
Db	1201	GAGATCGTGTATCTTACAGTACATGGAGCAGCTGTAGCTGGGCAAGCAGCTGGAGATCGGC	1260
QY	1261	CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTTCACACC	1320
Db	1261	CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTTCACACC	1320
QY	1321	CCCGACAAGAAGCACAGAAAGGAGCCCCCTTCTGTGTGATGGGCTTACGAGCTGACCCCC	1380
Db	1321	CCCGACAAGAAGCACAGAAAGGAGCCCCCTTCTGTGTGATGGGCTTACGAGCTGACCCCC	1380
QY	1381	GACAACTGAGCGTGCAGGCCATCGAGCTGCGCGAGAAAGGAGAGCTGGACCCGTGAACGAC	1440
Db	1381	GACAACTGAGCGTGCAGGCCATCGAGCTGCGCGAGAAAGGAGAGCTGGACCCGTGAACGAC	1440
QY	1441	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG	1500
Db	1441	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAAGGTG	1500
QY	1501	CGCCAGCTGTGCAAGCTGTGCGCGCCCAAGGCCCTTGACCGACATGCTGTGCCCTGACC	1560
Db	1501	CGCCAGCTGTGCAAGCTGTGCGCGCCCAAGGCCCTTGACCGACATGCTGTGCCCTGACC	1560
QY	1561	GAGGAGCGAGCTGAGCTGCGCGGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC	1620
Db	1561	GAGGAGCGAGCTGAGCTGCGCGGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGC	1620
QY	1621	GTGTACTTACGACCCAGCAAGGACCTGTGTGGCGCGAGATCCAGAAAGCAGGCGCACGACCAG	1680
Db	1621	GTGTACTTACGACCCAGCAAGGACCTGTGTGGCGCGAGATCCAGAAAGCAGGCGCACGACCAG	1680
QY	1681	TGGACCTTACAGATCTTACAGGAGCCCTTCAAGAACCTGGAAGACCCGCGCAAGTACGCCAAG	1740
Db	1681	TGGACCTTACAGATCTTACAGGAGCCCTTCAAGAACCTGGAAGACCCGCGCAAGTACGCCAAG	1740
QY	1741	ATGCGCAGCGCCACACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCC	1800
Db	1741	ATGCGCAGCGCCACACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCC	1800
QY	1801	ATGCGAGCATCTGTGATCTTGGGCAAGACCCCAAGTTCGCGCTGCCCATCAGAAAGGAG	1860
Db	1801	ATGCGAGCATCTGTGATCTTGGGCAAGACCCCAAGTTCGCGCTGCCCATCAGAAAGGAG	1860
QY	1861	ACCTGGGAGACTGTGTGACCGACTACTGTGAGGCCACCTGTGATCCCGAGTGGGAGTTC	1920
Db	1861	ACCTGGGAGACTGTGTGACCGACTACTGTGAGGCCACCTGTGATCCCGAGTGGGAGTTC	1920
QY	1921	GTGAAACACCCCGCTGTGTGAGCTGTGGTACCACTGAGGAGGAGCCCATCATCGGC	1980
Db	1921	GTGAAACACCCCGCTGTGTGAGCTGTGGTACCACTGAGGAGGAGCCCATCATCGGC	1980
QY	1981	GCCGAGACTTCTACCTGGAACGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGCGC	2040
Db	1981	GCCGAGACTTCTACCTGGAACGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGCGC	2040
QY	2041	TACGTGACCCGACCGCGGCGCGGAGAGTGTGAGCTTGACCGAGACCAACCAAGGAG	2100
Db	2041	TACGTGACCCGACCGCGGCGCGGAGAGTGTGAGCTTGACCGAGACCAACCAAGGAG	2100
QY	2101	ACCGAGCTGAGGCCATTCAGCTGGGCCCTGCGAGACAGCGGCGAGGCTGAACATCGTG	2160
Db	2101	ACCGAGCTGAGGCCATTCAGCTGGGCCCTGCGAGACAGCGGCGAGGCTGAACATCGTG	2160
QY	2161	ACCGACAGCAGTACCGCTTGGGCATCTACCGCGCCAGCCCGACAAAGAGCGAGGAGGAG	2220
Db	2161	ACCGACAGCAGTACCGCTTGGGCATCTACCGCGCCAGCCCGACAAAGAGCGAGGAGGAG	2220
QY	2221	CTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACTTGTAGCTGGGTG	2280
Db	2221	CTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACTTGTAGCTGGGTG	2280

Db 2221 CTGGTGAACAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACCTGAGCTGGGTG 2280  
Qy 2281 CCCGCCCAAGAGGATCGCGGCAACAGAGCAGATCGAAGCTGGTGAGCAAGGGCATC 2340  
Db 2281 CCCGCCCAAGAGGATCGCGGCAACAGAGCAGATCGAAGCTGGTGAGCAAGGGCATC 2340  
Qy 2341 CGCAAGTGTCTTCTCGAGCGGATCGATCGCGGCATCGTGACTACCAAGTACATGAC 2400  
Db 2341 CGCAAGTGTCTTCTCGAGCGGATCGATCGCGGCATCGTGACTACCAAGTACATGAC 2400  
Qy 2401 GACCTGTACGTGGCGAGCGCGGCCCTAGGATCGAATTAAGCTTCCCGGGCTAGCAC 2460  
Db 2401 GACCTGTACGTGGCGAGCGCGGCCCTAGGATCGAATTAAGCTTCCCGGGCTAGCAC 2460  
Qy 2461 GGTGAATTC 2469  
Db 2461 GGTGAATTC 2469

RESULT 2  
US-10-190-435-45  
; Sequence 45, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MESEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190.435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Poliopt\_C  
US-10-190-435-45

Query Match 99.5%; Score 2457; DB 15; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 GCCACCATGGCCGAGGCGCATGAGCAGGCCACAGCGGCCAACATCTCTGATGCGCGCAGC 66  
Db 1 GCCACCATGGCCGAGGCGCATGAGCAGGCCACAGCGGCCAACATCTCTGATGCGCGCAGC 60  
Qy 67 AACTTCAAGGSCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATC 126  
Db 61 AACTTCAAGGSCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATC 120  
Qy 127 GCCCGCAACTGCCCGCCCCCGCCGCAAGAGGGCTGCTGGAGTGGCGGCAAGAGGGCCAC 186  
Db 121 GCCCGCAACTGCCCGCCCCCGCCGCAAGAGGGCTGCTGGAGTGGCGGCAAGAGGGCCAC 180  
Qy 187 CAGATGAAGAGTGCACCGAGCGCGCAGGCCCAACTTCTTCGCGAGGAGACTGGCCCTTCCC 246  
Db 181 CAGATGAAGAGTGCACCGAGCGCGCAGGCCCAACTTCTTCGCGAGGAGACTGGCCCTTCCC 240  
Qy 247 CAGGGCAAGCGCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCACCGAGCGGC 306  
Db 241 CAGGGCAAGCGCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCACCGAGCGGC 300  
Qy 307 GAGCTGAGGTGCGCGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCAC 366  
Db 301 GAGCTGAGGTGCGCGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCAC 360  
Qy 367 CTGAACCTTCCCCCAGATCAACCTGTGGAGCGCGCCCCCTGTGTAGCATCAAGGTGGCGGC 426

Db 361 CTGAACCTTCCCCCAGATCAACCTGTGTGAGCGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 420  
Qy 427 CAGATCAAGAGGCGCTGTGTGAGCACCGGCGCGCAGACACCCGCTGTGTGAGGAGATGAGC 486  
Db 421 CAGATCAAGAGGCGCTGTGTGAGCACCGGCGCGCAGACACCCGCTGTGTGAGGAGATGAGC 480  
Qy 487 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGCGCTTTCATCAAGGTGGCG 546  
Db 481 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGCGCTTTCATCAAGGTGGCG 540  
Qy 547 CAGTACCAACAGATCTCTGATCGAGATCTGCGGCAAGAGCCATCGCGCACCGTCTCTGATC 606  
Db 541 CAGTACCAACAGATCTCTGATCGAGATCTGCGGCAAGAGCCATCGCGCACCGTCTCTGATC 600  
Qy 607 GGCCCCACCCCCCGTGAACATCATCGCGCGCAACATGCTGTGACCCAGCTGGGCTGCACCTG 666  
Db 601 GGCCCCACCCCCCGTGAACATCATCGCGCGCAACATGCTGTGACCCAGCTGGGCTGCACCTG 660  
Qy 667 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGAGCGGC 726  
Db 661 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGAGCGGC 720  
Qy 727 CCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
Db 721 CCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 780  
Qy 787 GAGGAGATGGAGAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAACC 846  
Db 781 GAGGAGATGGAGAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAACC 840  
Qy 847 CCGGTGTTCCCATCAAGAAAGAGGACAGCACCAAGTGGCGGCAAGCTGTGGAGATCTCGC 906  
Db 841 CCGGTGTTCCCATCAAGAAAGAGGACAGCACCAAGTGGCGGCAAGCTGTGGAGATCTCGC 900  
Qy 907 GAGTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGCC 966  
Db 901 GAGTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGCC 960  
Qy 967 GGCCTGAAGAAAGAGAGCGGTGACCGTCTGACCGTGGCGCGACCGCTACTTTCAGCGGTG 1026  
Db 961 GGCCTGAAGAAAGAGAGCGGTGACCGTCTGAGCGTGGCGGACCGCTACTTTCAGCGGTG 1020  
Qy 1027 CCGCTGAGCAGGACTTCCCGCAAGTACACCGCTTTCACCATCCCGCAGATCAACAACAG 1086  
Db 1021 CCGCTGAGCAGGACTTCCCGCAAGTACACCGCTTTCACCATCCCGCAGATCAACAACAG 1080  
Qy 1087 ACCCGCGCATCCGCTACCAAGTACACCGTCTGCCCGGCTGGAGGGCAGCCCCCAGC 1146  
Db 1081 ACCCGCGCATCCGCTACCAAGTACACCGTCTGCCCGGCTGGAGGGCAGCCCCCAGC 1140  
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGTGAGCGCTTTCGCGCGCCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGTGAGCGCTTTCGCGCGCCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACAGTACATGAGACCTGTGACGAGCGACCTGTGAGATCGGCGCAGAC 1266  
Db 1201 GTGATCTACAGTACATGAGACCTGTGACGAGCGACCTGTGAGATCGGCGCAGAC 1260  
Qy 1267 CCGCGCAAGATCGAGGAGCTGCGCAAGCATCTGTGCGCTGGGGCTTTCACACCCCGGAC 1326  
Db 1261 CCGCGCAAGATCGAGGAGCTGCGCAAGCATCTGTGCGCTGGGGCTTTCACACCCCGGAC 1320  
Qy 1327 AAGAGCACCAAGAGAGCGCCCCCTTCTGTGTGATGGGCTACGAGCTGCAACCCCGCAAG 1386  
Db 1321 AAGAGCACCAAGAGAGCGCCCCCTTCTGTGTGATGGGCTACGAGCTGCAACCCCGCAAG 1380  
Qy 1387 TGAACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1446  
Db 1381 TGAACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1440  
Qy 1447 AAGCTGTGGGCAAGCTGAACCTGTGGCGCAGCATCTACCCCGGATCAAGGTGGCGGCAG 1506

```

1441 AAGTCGTGGCGAAGCTGAACCTGGCCGAGCCAGATCTACCCCGCATCAAGGTGCGCCAG 1500
1507 CTGTGAAGCTGTGCGCGCGCCAGCCCTGACGACATCTGCCCCTGACCGAGGAG 1566
1501 CTGTGAAGCTGTGCGCGCGCCAGCCCTGACGACATCTGCCCCTGACCGAGGAG 1560
1567 GCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGACGCGCGTGAC 1626
1561 GCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGACGCGCGTGAC 1620
1627 TAGACCCCAAGCAAGCACTGGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAGTGAC 1686
1621 TAGACCCCAAGCAAGCACTGGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAGTGAC 1680
1687 TACCAGATCTACGAGGAGCCCTTCAGAACTGAGACCGGCAAGTACGCGCAAGATGCGC 1746
1681 TACCAGATCTACGAGGAGCCCTTCAGAACTGAGACCGGCAAGTACGCGCAAGATGCGC 1740
1747 ACCGCCACACCAAGCAAGCTGAAGCAGCTGACCGAGCCGTCAGAAAGATCGCCATGGAG 1806
1741 ACCGCCACACCAAGCAAGCTGAAGCAGCTGACCGAGCCGTCAGAAAGATCGCCATGGAG 1800
1807 AGCATCGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCCATCCAGAAGAGGACCTGG 1866
1801 AGCATCGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCGCCATCCAGAAGAGGACCTGG 1860
1867 GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1926
1861 GAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAAC 1920
1927 ACCCCCCCTGGTGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCTGCGCGCCGAG 1986
1921 ACCCCCCCTGGTGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCTGCGCGCCGAG 1980
1987 ACCTTCTACGTGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2046
1981 ACCTTCTACGTGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
2047 ACCGACCGGGCGCGGAGAGATGCTGAGCTTACCGAGACCAACCAAGAGAGACCGAG 2106
2041 ACCGACCGGGCGCGGAGAGATGCTGAGCTTACCGAGACCAACCAAGAGAGACCGAG 2100
2107 CTGACGCGCATCAGCTGCGCCCTGACGACAGCGGCGAGAGGTGAACATCGTGACCGAC 2166
2101 CTGACGCGCATCAGCTGCGCCCTGACGACAGCGGCGAGAGGTGAACATCGTGACCGAC 2160
2167 AGCCAGTACCGCTTGGSCATCATCCAGGCCAGCCCGACAAAGACGAGAGCGAGCTGGTG 2226
2161 AGCCAGTACCGCTTGGSCATCATCCAGGCCAGCCCGACAAAGACGAGAGCGAGCTGGTG 2220
2227 AACGAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCGCC 2286
2221 AACGAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCGCC 2280
2287 CACAAGGCGCATCGCGGCAACAGCAGATCGAAGCTGTGAGCAAGGCGCATCCGCAAG 2346
2281 CACAAGGCGCATCGCGGCAACAGCAGATCGAAGCTGTGAGCAAGGCGCATCCGCAAG 2340
2347 GTGCTGTTCTTGGACCGCATCGATGGCGGATCGTGATCTACAGTACATGAGACGACCTG 2406
2341 GTGCTGTTCTTGGACCGCATCGATGGCGGATCGTGATCTACAGTACATGAGACGACCTG 2400
2407 TAGTGGGCGAGCGCGCCCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2463
2401 TAGTGGGCGAGCGCGCCCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

```

RESULT 3

```

US-10-190-305A-39
; Sequence 39, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan

```

```

; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT FILING DATE: 2002-07-05
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-305A-39

```

```

Query Match 99.5%; Score 2457; DB 16; Length 2457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCACCATGCGCGAGGCCCATGAGCCAGGCGCCACGAGCCCAACATCTGATGCGAGCGCAGC 66
DB 1 GCCACCATGCGCGAGGCCCATGAGCCAGGCGCCACGAGCCCAACATCTGATGCGAGCGCAGC 60
QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATC 126
DB 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATC 120
QY 127 GCCCGCAACTGCGCGCGCCCCCGCCCAAGAGGGGTGTGTGAAGTGTGCGGCAAGAGGGGCCAC 186
DB 121 GCCCGCAACTGCGCGCGCCCCCGCCCAAGAGGGGTGTGTGAAGTGTGCGGCAAGAGGGGCCAC 180
QY 187 CAGATGAAGGACTGCAACGAGCGCCAGGCGCAACTTCTTCCGAGAGACCTGGGCTTCCCC 246
DB 181 CAGATGAAGGACTGCAACGAGCGCCAGGCGCAACTTCTTCCGAGAGACCTGGGCTTCCCC 240
QY 247 CAGGGCAAGGCGCGGAGTTTCCCGAGCGAGCAGAAACCGGCGCCAAAGCCCCACAGCCGCGC 306
DB 241 CAGGGCAAGGCGCGGAGTTTCCCGAGCGAGCAGAAACCGGCGCCAAAGCCCCACAGCCGCGC 300
QY 307 GAGCTCAGGTGCGCGCGCGCAACACCCCGCAGCGGCGCGGCCCGCCGAGCGCCAGGGGACC 366
DB 301 GAGCTCAGGTGCGCGCGCGCAACACCCCGCAGCGGCGCGGCCCGCCGAGCGCCAGGGGACC 360
QY 367 CTGAACTTCCCCCAGATCACTCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGGCGGC 426
DB 361 CTGAACTTCCCCCAGATCACTCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGGCGGC 420
QY 427 CAGATCAAGGAGGCGCTGTCTGGACACCGCGCGCGCAGACACCGTGTGGAGGAGATGAGC 486
DB 421 CAGATCAAGGAGGCGCTGTCTGGACACCGCGCGCGCAGACACCGTGTGGAGGAGATGAGC 480
QY 487 CTGCCCGGCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTTCATCAAGGTGCGC 546
DB 481 CTGCCCGGCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTTCATCAAGGTGCGC 540
QY 547 CAGTACGACGAGATCTGTATCGAGATCTGGGCAAGAGCCCATCGSCACCGTGTGATC 606
DB 541 CAGTACGACGAGATCTGTATCGAGATCTGGGCAAGAGCCCATCGSCACCGTGTGATC 600
QY 607 GGCCCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGGTGGGCTGCACCCCTG 666
DB 601 GGCCCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGGTGGGCTGCACCCCTG 660
QY 667 AACTTCCCCATCAGCCCCATTCGAGACCGTGTCCCGTGAAGCTGAAGCCCGGCGATGAGCGGC 726
DB 661 AACTTCCCCATCAGCCCCATTCGAGACCGTGTCCCGTGAAGCTGAAGCCCGGCGATGAGCGGC 720
QY 727 CCCAAGGTGAAGCAGTGGGCCCCCTGACCGGAGGAGAAATCAAGGCCCTGACCGCCCATCTGC 786
DB 721 CCCAAGGTGAAGCAGTGGGCCCCCTGACCGGAGGAGAAATCAAGGCCCTGACCGCCCATCTGC 780

```



QY	61	CGCAGCAACTTTCAAGSGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCG	120
Db	61	CGCAGCAACTTTCAAGSGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCG	120
QY	121	CACATCGCCGCAACTGCGCGCCCCCGCAGAGGGTGTCTGGAAGTGCGGCAAGGAG	180
Db	121	CACATCGCCGCAACTGCGCGCCCCCGCAGAGGGTGTCTGGAAGTGCGGCAAGGAG	180
QY	181	GGCCACCATGAAGNACTGCACGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGCC	240
Db	181	GGCCACCATGAAGNACTGCACGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGCC	240
QY	241	TTCCCCCAGGCAAGGCGCCGAGTTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Db	241	TTCCCCCAGGCAAGGCGCCGAGTTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
QY	301	AGCGCGAGTGCAGGTGCGCGCGCAACCCCCCGCAGCGAGGCCCGCGCGAGCCCCAG	360
Db	301	AGCGCGAGTGCAGGTGCGCGCGCAACCCCCCGCAGCGAGGCCCGCGCGAGCCCCAG	360
QY	361	GGCACCTGAACTTCCCCCAGATCACTCTGTGCGAGGCCCTTGTGTAGCATCAAGGTG	420
Db	361	GGCACCTGAACTTCCCCCAGATCACTCTGTGCGAGGCCCTTGTGTAGCATCAAGGTG	420
QY	421	GGCGGCCAGATCAAGGAGGCCCTGTGGACACCGGCGCGACGACACCGTGTGGAGGAG	480
Db	421	GGCGGCCAGATCAAGGAGGCCCTGTGGACACCGGCGCGACGACACCGTGTGGAGGAG	480
QY	481	ATGAGCCTTGC CGCGCAAGTGGAAAGCCCCAAGATGATCGGGGCATCGCGCGCTTTCATCAAG	540
Db	481	ATGAGCCTTGC CGCGCAAGTGGAAAGCCCCAAGATGATCGGGGCATCGCGCGCTTTCATCAAG	540
QY	541	GTGCGCCAGTACGACAGATCTCTGTAGATCTGCGGCAAGAGGGCATCGGCACCGTG	600
Db	541	GTGCGCCAGTACGACAGATCTCTGTAGATCTGCGGCAAGAGGGCATCGGCACCGTG	600
QY	601	CTGATCGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC	660
QY	661	ACCCTGAACTTCCCCATCGAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGCATG	720
Db	661	ACCCTGAACTTCCCCATCGAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGCATG	720
QY	721	GACGGCCCCAAGTGAAGCATGAGGAGGCAAGATCAACAAGATCGGCCCGAGAACCCCTAC	780
Db	721	GACGGCCCCAAGTGAAGCATGAGGAGGCAAGATCAACAAGATCGGCCCGAGAACCCCTAC	780
QY	841	AACACCCCGTGTTCGCCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Db	841	AACACCCCGTGTTCGCCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC	900
QY	901	TTCCGGAGCTGAACAAAGCCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCAC	960
Db	901	TTCCGGAGCTGAACAAAGCCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCAC	960
QY	961	CCCGCGGCTGAAAGAGAGAGAGCGTGACCTGTGACCTGGCGACCGCTTACTTC	1020
Db	961	CCCGCGGCTGAAAGAGAGAGAGCGTGACCTGTGACCTGGCGACCGCTTACTTC	1020
QY	1021	AGCGTGCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC	1080
Db	1021	AGCGTGCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC	1080
QY	1081	AACGAGACCCCGGCATCCGCTACAGTACAAAGTGTGCTGCCCGCAGGGCTGGAGGGGAGC	1140
Db	1081	AACGAGACCCCGGCATCCGCTACAGTACAAAGTGTGCTGCCCGCAGGGCTGGAGGGGAGC	1140

QY	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCGCAACCCC	1200
Db	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCGCAACCCC	1200
QY	1201	GAGATCGTGATCTACAGTACATGAGACCTGTAGCTGGGCAAGCACTTGGAGATCGGC	1260
Db	1201	GAGATCGTGATCTACCA-----GGCCCGCTGTAGCTGGGCAAGCACTTGGAGATCGGC	1254
QY	1261	CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGCTTTCACCAAC	1320
Db	1255	CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGCTTTCACCAAC	1314
QY	1321	CCCGACAAGAAGCACCAAGAAGGAGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCC	1380
Db	1315	CCCGACAAGAAGCACCAAGAAGGAGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCC	1374
QY	1381	GACAACTGACACCGTGCAGCCCATCGAGCTGCGCGAAGAGAGCTTGAACCGTGAACGAC	1440
Db	1375	GACAACTGACACCGTGCAGCCCATCGAGCTGCGCGAAGAGAGCTTGAACCGTGAACGAC	1434
QY	1441	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1500
Db	1435	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1494
QY	1501	CGCAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTTGACCAATCTGTGCCCCCTGACC	1560
Db	1495	CGCAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTTGACCAATCTGTGCCCCCTGACC	1554
QY	1561	GAGGAGCGAGCTGAGCTGCGCGAGAACCGGAGATCTCTGCGCGAGCCCGTGCACGGC	1620
Db	1555	GAGGAGCGAGCTGAGCTGCGCGAGAACCGGAGATCTCTGCGCGAGCCCGTGCACGGC	1614
QY	1621	GTGTACTACGACCCCAAGAGGACCTGTGTGGCGGAGATCCAGAAGCAGGGCCACGACCAG	1680
Db	1615	GTGTACTACGACCCCAAGAGGACCTGTGTGGCGGAGATCCAGAAGCAGGGCCACGACCAG	1674
QY	1681	TGGACCTTACAGATCTTACCAAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCCAAG	1740
Db	1675	TGGACCTTACAGATCTTACCAAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTACGCCAAG	1734
QY	1741	ATGCGCACCGCCCAACCAAGACGTGAAGCAGCTGACGAGGCCGTGCAGAAAGTTCGCC	1800
Db	1735	ATGCGCACCGCCCAACCAAGACGTGAAGCAGCTGACGAGGCCGTGCAGAAAGTTCGCC	1794
QY	1801	ATGGAAGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAG	1860
Db	1795	ATGGAAGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAG	1854
QY	1861	ACCTGGGAGACCTGTGTGACCGACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTC	1920
Db	1855	ACCTGGGAGACCTGTGTGACCGACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTC	1914
QY	1921	GTGAAACACCCCGCTGTGTGAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGC	1980
Db	1915	GTGAAACACCCCGCTGTGTGAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGC	1974
QY	1981	GCGGAGACCTTCTACCTTGAACCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGCGC	2040
Db	1975	GCGGAGACCTTCTACCTTGAACCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGCGC	2034
QY	2041	TACGTGACCGAGCCGGGCGCGCAAGATCTGTGAGCTTGAACCGAGACCAACCAAGAG	2100
Db	2035	TACGTGACCGAGCCGGGCGCGCAAGATCTGTGAGCTTGAACCGAGACCAACCAAGAG	2094
QY	2101	ACCGAGCTGAGGCCATTCAGCTGGGCCCTTGACGAGCAGCGGCGAGGCTGAACATCGTG	2160
Db	2095	ACCGAGCTGAGGCCATTCAGCTGGGCCCTTGACGAGCAGCGGCGAGGCTGAACATCGTG	2154
QY	2161	ACCGAGCAGCAGTACCGCTTGGGCATCATCCAGGCCCGAGCCGACAAAGAGCGAGCGAG	2220
Db	2155	ACCGAGCAGCAGTACCGCTTGGGCATCATCCAGGCCCGAGCCCGACAAAGAGCGAGCGAG	2214
QY	2221	CTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAGGAGAAAGGTGTACTCTGAGCTGGGTG	2280

```
Db 2215 |||||CTGTGAACCAAGATCATCGACAGCTGATCAAGAGGAGAGAGGTGTACTCTGAGCTGGGTG 2274
Qy 2281 CCGGCCCAACAAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db 2275 CCGGCCCAACAAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATC 2334
Qy 2341 CGCAAGTGCTGTTCTCTGGACGGCATCGATCGCGGGCATCTGTGATCTACCAGTACATGGAC 2400
Db 2335 CGCAAGTGCTGTTCTCTGGACGGCATCGATCGCGGGCATCTGTGATCTACCAGTACATGGAC 2394
Qy 2401 GACCTGTACGTGGCAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Db 2395 GACCTGTACGTGGCAGCGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2454
Qy 2461 GGTGAATTC 2469
Db 2455 GGTGAATTC 2463
```

## RESULT 5

```
US-10-190-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-435-44
```

```
Query Match 98.6%; Score 2434.6; DB 15; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 GTCGACGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCGAGCGCCAAACATCCTGATGCAG 60
Db 1 GTCGACGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCGAGCGCCAAACATCCTGATGCAG 60

Qy 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGGGC 120
Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGGGC 120

Qy 121 CACATCCCGCGCAACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180
Db 121 CACATCCCGCGCAACTGCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180

Qy 181 GGCCACCAAGTGAAGAGTGCACCGAGCGCCCAACCTTCTTCCGCGAGGACCTGGCC 240
Db 181 GGCCACCAAGTGAAGAGTGCACCGAGCGCCCAACCTTCTTCCGCGAGGACCTGGCC 240

Qy 241 TTCCCCCAGGCAAGGCCCGGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300
Db 241 TTCCCCCAGGCAAGGCCCGGAGTTCCCCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300

Qy 301 AGCGCGAGCTGCAGGTGCGCGGACAAACCCCGCAGCGAGGCGCGCGCGCGCGAG 360
Db 301 AGCGCGAGCTGCAGGTGCGCGGACAAACCCCGCAGCGAGGCGCGCGCGCGCGAG 360
```

```
Qy 361 GGCAACCTGAACCTTCCCCCAGATCACTCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTG 420
Db 361 GGCAACCTGAACCTTCCCCCAGATCACTCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTG 420

Qy 421 GGCGGCCAGATCAAGGAGGCCCTCTGTGACACCGCGCGCCGACGACACCGTGTCTGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCCCTCTGTGACACCGCGCGCCGACGACACCGTGTCTGAGGAG 480

Qy 481 ATGAGCTGTCGCCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCTGGCGGCTTTCATCAAG 540
Db 481 ATGAGCTGTCGCCGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCTGGCGGCTTTCATCAAG 540

Qy 541 GTGCGCCAGTACGACAGATCTCTGATCGAGATCTCGCGCAGAAAGGCGCATCGGCACCGTG 600
Db 541 GTGCGCCAGTACGACAGATCTCTGATCGAGATCTCGCGCAGAAAGGCGCATCGGCACCGTG 600

Qy 601 CTGATCGGCCCCACCGCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCCCACCGCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660

Qy 661 ACCCTGAATTCCTCCCATCAGCCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGATG 720
Db 661 ACCCTGAATTCCTCCCATCAGCCCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGATG 720

Qy 721 GACGCCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCTTACCGCC 780
Db 721 GACGCCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCTTACCGCC 780

Qy 781 ATCTGCGAGGAGATGGAGAGGAGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTTAC 840
Db 781 ATCTGCGAGGAGATGGAGAGGAGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTTAC 840

Qy 841 AACACCCCCCTGTTTCGCCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTTGTGGAC 900
Db 841 AACACCCCCCTGTTTCGCCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTTGTGGAC 900

Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTTGGGAGGTGAGCTGGGATCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTTGGGAGGTGAGCTGGGATCCCCAC 960

Qy 961 CCGCGCGGCTGAAGAAGAAGAGAGCGTGACCGTGTGACGCGTGGCGAGCGCTACTTTC 1020
Db 961 CCGCGCGGCTGAAGAAGAAGAGAGCGTGACCGTGTGACGCGTGGCGAGCGCTACTTTC 1020

Qy 1021 AGCGTGCCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCCAGCATCAAC 1080

Qy 1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAAACGCTGCTGCGGCGAGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAAACGCTGCTGCGGCGAGGCTGGAAGGGCAGC 1140

Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCGCTTCCGCGCGCCGCAACCCC 1200
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCGCTTCCGCGCGCCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTACCAAGTACATGGACGACTGTGTGCTGGGCGAGCGACTTGGAGATCGGC 1260
Db 1201 GAGATCGTGATCTACCAAGTACATGGACGACTGTGTGCTGGGCGAGCGACTTGGAGATCGGC 1260

Qy 1261 CAGCAGCCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCTGGGCTTTCACAC 1320
Db 1261 CAGCAGCCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGTGCTGGGCTTTCACAC 1320

Qy 1321 CCCGACAAGAGCAGCAGAGAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGACACCC 1380
Db 1321 CCCGACAAGAGCAGCAGAGAGGAGCCCCCTTCTGTGGATGGGCTACGAGCTGACACCC 1380

Qy 1381 GACAAGTGAACCGTGCAGCGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1440
Db 1381 GACAAGTGAACCGTGCAGCGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1440

Qy 1441 ATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGAGCGAGATCTACCCCGGATCAAGGTG 1500
```





QY 721 GACGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAAGATCAAGGCCCTGACCGCC 780  
Db 721 GACGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAAGATCAAGGCCCTGACCGCC 780  
QY 781 ATCTGCGAGGAGATGGAGAGGCGCAAGATCAACCAAGATGGCGCCCGCAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAGAGGCGCAAGATCAACCAAGATGGCGCCCGCAACCCCTAC 840  
QY 841 AACACCCCGCTGTTCCGCATCAAGAAGAACAGACACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGCTGTTCCGCATCAAGAAGAACAGACACCAAGTGGCGCAAGCTGGTGGAC 900  
QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGCACTTCTGGAGAGTGCAAGTGGGCAATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGCACTTCTGGAGAGTGCAAGTGGGCAATCCCCAC 960  
QY 961 CCCGCGCCCTGAAGAGAGAGAGCGTGAACCGTGTGAGCGTGGCGAGCGCTACTTC 1020  
Db 961 CCCGCGCCCTGAAGAGAGAGAGCGTGAACCGTGTGAGCGTGGCGAGCGCTACTTC 1020  
QY 1021 AGCGTGGCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTTCACCATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGGCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTTCACCATCCCCAGCATCAAC 1080  
QY 1081 AACGAGACCCCGCGCATCCGCTACAGTACAACTGTGCTGCCCGCAGAGGCTGGAGGCGAGC 1140  
Db 1081 AACGAGACCCCGCGCATCCGCTACAGTACAACTGTGCTGCCCGCAGAGGCTGGAGGCGAGC 1140  
QY 1141 CCCAGCATCTTCCAGAGCGAGATCAACCAAGATCTGGAGCCCTTCCGCGCCCGCAACCCC 1200  
Db 1141 CCCAGCATCTTCCAGAGCGAGATCAACCAAGATCTGGAGCCCTTCCGCGCCCGCAACCCC 1200  
QY 1201 GAGATCGTGATCTACCAAGTACATGAGACGACCTGTGAGCGAGCAGCTGGAGATCGGC 1260  
Db 1201 GAGATCGTGATCTACCA-----GGCCCCCTGTAGTGGGACGAGCTGGAGATCGGC 1260  
QY 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGTGCTGCGTGGGCTTCCACACC 1320  
Db 1261 CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACTGTGCTGCGTGGGCTTCCACACC 1320  
QY 1321 CCCGACAGAGACACCAAGAGCGCCCTTCTGTGATGGCTAGGAGCTGACGCC 1380  
Db 1321 CCCGACAGAGACACCAAGAGCGCCCTTCTGTGATGGCTAGGAGCTGACGCC 1380  
QY 1381 GACAAAGTGGACCGTGCAGCCCATCGAGTGCAGAGAGAGAGCTGACCGTGAACGAC 1440  
Db 1381 GACAAAGTGGACCGTGCAGCCCATCGAGTGCAGAGAGAGAGCTGACCGTGAACGAC 1440  
QY 1441 ATCCAGAAAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1500  
Db 1441 ATCCAGAAAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1500  
QY 1501 CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATGTGCGCCCTGACC 1560  
Db 1501 CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATGTGCGCCCTGACC 1560  
QY 1561 GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCGCAGATCTGCGCGAGCGCCGTGCACGGC 1620  
Db 1561 GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCGCAGATCTGCGCGAGCGCCGTGCACGGC 1620  
QY 1621 GTGTACTACGACCCAGCAAGGACTGTGGTGGCGCGAGATCCAGAAAGAGGCGCCAGCCAG 1680  
Db 1621 GTGTACTACGACCCAGCAAGGACTGTGGTGGCGCGAGATCCAGAAAGAGGCGCCAGCCAG 1680  
QY 1681 TGGACCTACCAAGTCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1740  
Db 1681 TGGACCTACCAAGTCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1740  
QY 1741 ATGCGCACCGCCACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGGAGAGATCGCC 1800  
Db 1741 ATGCGCACCGCCACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGGAGAGATCGCC 1800  
QY 1794 ATGCGCACCGCCACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGGAGAGATCGCC 1794

QY 1801 ATGAGAGCATCGTGTGATCTGGGGCAAGACCCCAAGTTTCCGCTGCCCATCCAGAAGGAG 1860  
Db 1795 ATGAGAGCATCGTGTGATCTGGGGCAAGACCCCAAGTTTCCGCTGCCCATCCAGAAGGAG 1854  
QY 1861 ACCTGGAGAGACTGTGGACCGACTACTTGGCAGGCGCACCTGATCCCGAGTGGGAGTTTC 1920  
Db 1855 ACCTGGAGAGACTGTGGACCGACTACTTGGCAGGCGCACCTGATCCCGAGTGGGAGTTTC 1914  
QY 1921 GTGAACACCCCGCTGTGGTGAAGCTGTGTGTACAGCTGGAGAGAGGCCCATCATCGGC 1980  
Db 1915 GTGAACACCCCGCTGTGGTGAAGCTGTGTGTACAGCTGGAGAGAGGCCCATCATCGGC 1974  
QY 1981 GCCGAGACCTTCTAGTGGACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2040  
Db 1975 GCCGAGACCTTCTAGTGGACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGCGC 2034  
QY 2041 TACGTGACCGACCGGGCGCGGCAAGATCTGTGAGCCTGTGAGCCTGTGAGACCGACCAACCAAG 2100  
Db 2035 TACGTGACCGACCGGGCGCGGCAAGATCTGTGAGCCTGTGAGCCTGTGAGACCGACCAACCAAG 2094  
QY 2101 ACCGAGCTGCAGGCGCATCTCAGCTGGCCCTGTCAGGACACGCGCAGCGAGGTGAACATCGTG 2160  
Db 2095 ACCGAGCTGCAGGCGCATCTCAGCTGGCCCTGTCAGGACACGCGCAGCGAGGTGAACATCGTG 2154  
QY 2161 ACCGAGACCGATGACGCGCTGGGCGATCATCAGGCCCGACCGAGCGAGCGAGCGAG 2220  
Db 2155 ACCGAGACCGATGACGCGCTGGGCGATCATCAGGCCCGACCGAGCGAGCGAGCGAG 2214  
QY 2221 CTGTGTGAACCAAGCATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTCAGCTGGGTG 2280  
Db 2215 CTGTGTGAACCAAGCATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTCAGCTGGGTG 2274  
QY 2281 CCCGCCACCAAGGCGCATCGCGCGCAAGCAGATCGACAGCTGGTGGAGCAAGGCGCATC 2340  
Db 2275 CCCGCCACCAAGGCGCATCGCGCGCAAGCAGATCGACAGCTGGTGGAGCAAGGCGCATC 2334  
QY 2341 CGCAAGTGTCTTCTTGGAGCGGATCGATCGCGCGCATCTGATCTTACAGTACATGAGAC 2400  
Db 2335 CGCAAGTGTCTTCTTGGAGCGGATCGATCGCGCGCATCTGATCTTACAGTACATGAGAC 2394  
QY 2401 GACCTGTACGTGGGCGCGCGCTAGGATCCATTAAAGCTTCCCGGGGCTAGCACC 2460  
Db 2395 GACCTGTACGTGGGCGCGCGCTAGGATCCATTAAAGCTTCCCGGGGCTAGCACC 2454  
QY 2461 GGT 2463  
Db 2455 GGT 2457

## RESULT 7

US-09-899-575-32  
; Sequence 32, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Batrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; CURRENT FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475,704  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PR975YMM

US-09-899-575-32

Query Match									
Best Local Similarity 97.8%; Score 2415.4; DB 10; Length 2457;									
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;									
Qy	1	GTCACGCGCACATGCGCGAGCCCATGAGCCAGGCGCCACCGAGCGCAATCTCTGATCGAG	60						
Db	1	GTCAGCGCCACATGCGCGAGCCCATGAGCGAGCGCACCGAGCGCAATCTCTGATCGAG	60						
Qy	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTCTTTAACTCGGCAAGGAGGGC	120						
Db	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTCTTTAACTCGGCAAGGAGGGC	120						
Qy	121	CACATCGCGCGCACTCGCGCGCCCCCGCAAGAGGGCTCTCTGGAGTGGCGCAAGGAG	180						
Db	121	CACATCGCGCGCAACTCGCGCGCCCCCGCAAGAGGGCTCTCTGGAGTGGCGCAAGGAG	180						
Qy	181	GGCCACCAGATGAAGGACTGCAACGAGCGCCAGCGCCAACTTTCTTCGCGAGGACTCGGCC	240						
Db	181	GGCCACCAGATGAAGGACTGCAACGAGCGCCAGCGCCAACTTTCTTCGCGAGGACTCGGCC	240						
Qy	241	TTCCGCCAGGGCAAGGCCCGGAGTTTCCGAGCGAGCAAGACCGCGCCCAACAGCCCCACC	300						
Db	241	TTCCGCCAGGGCAAGGGCCCGGAGTTTCCCGAGCGAGCAAGACCGCGCCCAACAGCCCCACC	300						
Qy	301	AGCCGCGAGCTGCAAGTTCGCGGCGCAACACCCCGCAGCGAGGCGCGCGAGCGCCAG	360						
Db	301	AGCCGCGAGCTGCAAGTTCGCGGCGCAACACCCCGCAGCGAGGCGCGCGAGCGCCAG	360						
Qy	361	GGCACCCTGAACTTTCCCGAGATCACCTGTGGAGCGCCCCCTTGGTAGCATCAAGGTG	420						
Db	361	GGCACCCTGAACTTTCCCGAGATCACCTGTGGAGCGCCCCCTTGGTAGCATCAAGGTG	420						
Qy	421	GGCGGCGAGTCAAGGAGGCCCTCTCTGGACACCGCGCGCAGCAGACCGTGTGGAGGAG	480						
Db	421	GGCGGCGAGTCAAGGAGGCCCTCTCTGGACACCGCGCGCAGCAGACCGTGTGGAGGAG	480						
Qy	481	ATGAGCCTGCGCGCAAGTGGAAAGCCAAAGATGATCGGCGGCATCTGGCGGTTCATCAAG	540						
Db	481	ATGAGCCTGCGCGCAAGTGGAAAGCCAAAGATGATCGGCGGCATCTGGCGGTTCATCAAG	540						
Qy	541	GTCGCGAGTACGACCGAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGACCGTG	600						
Db	541	GTCGCGAGTACGACCGAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGACCGTG	600						
Qy	601	CTGATCGGCGCCACCCCGTGAAACATCATCGGCGCAACATCTGACCCAGCTGGCGTGC	660						
Db	601	CTGATCGGCGCCACCCCGTGAAACATCATCGGCGCAACATCTGACCCAGCTGGCGTGC	660						
Qy	661	ACCTGGAATTTCCCATCAGCCCCCATCGAGACCGTGTGCCGTGGAAGCTCGGCGATG	720						
Db	661	ACCTGGAATTTCCCATCAGCCCCCATCGAGACCGTGTGCCGTGGAAGCTCGGCGATG	720						
Qy	721	GACGGCCCCAGGTTGAGAGTGGCCCTTGACCGGAGGAGATCAAGGCCCTCAGCCGCC	780						
Db	721	GACGGCCCCAGGTTGAGAGTGGCCCTTGACCGGAGGAGATCAAGGCCCTCAGCCGCC	780						
Qy	781	ATCTGCGAGGATGGAGAGGGGCAAGATCAACAGATCGGCGCCCGAGAAACCCCTAC	840						
Db	781	ATCTGCGAGGATGGAGAGGGGCAAGATCAACAGATCGGCGCCCGAGAAACCCCTAC	840						
Qy	841	AACACCCCGTGTTCGCCATCAAGAGAGAGACACCAAGTGGCGGCAAGCTGGTGAC	900						
Db	841	AACACCCCGTGTTCGCCATCAAGAGAGAGACACCAAGTGGCGGCAAGCTGGTGAC	900						
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCAC	960						
Db	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCAC	960						
Qy	961	CCCGCCGCTTGAAGAGAAAGAGCGTGACCGTGTGGAAGTGGGCGACGCTACTTTC	1020						
Db	961	CCCGCCGCTTGAAGAGAAAGAGCGTGACCGTGTGGAAGTGGGCGACGCTACTTTC	1020						

```
Qy 2101 ACCGAGTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGCGAGGTTGAAATCGTG 2160
Db 2089 ACCGAGTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGCGAGGTTGAAATCGTG 2148
Qy 2161 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCGACCAAGAGCGAGCGAG 2220
Db 2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCGACCAAGAGCGAGCGAG 2208
Qy 2221 CTGCTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGTTACCTGAGCTGGTG 2280
Db 2209 CTGCTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGTTACCTGAGCTGGTG 2268
Qy 2281 CCCGCCACAAAGGCGATCGCGGCAACAGAGCAGATCGAAGCTGGTGAGCAAGGCGATC 2340
Db 2269 CCCGCCACAAAGGCGATCGCGGCAACAGAGCAGATCGAAGCTGGTGAGCAAGGCGATC 2328
Qy 2341 CGCAAGTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTATACGAC 2400
Db 2329 CGCAAGTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTATACGAC 2388
Qy 2401 GACCTGTACGTGGGACGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Db 2389 GACCTGTACGTGGGACGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2448
Qy 2461 GGTGAATTC 2469
Db 2449 GGTGAATTC 2457
```

## RESULT 8

```
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MECEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBERG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMMW_C
US-10-190-435-43
```

```
Query Match 97.3%; Score 2401.8; DB 15; Length 2445;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;
```

```
Qy 7 GCCACCATGGCCGAGGCGCATGAGCCAGGCCAGCCAGCCCAACATCTTGATGCGCGCAGC 66
Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGGCCAGCCAGCCCAACATCTTGATGCGCGCAGC 60
Qy 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 126
Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGGAGGGCCACATC 120
Qy 127 GCCCGCAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGGGAAGTGGCGCAAGAGGGGCCAC 186
Db 121 GCCCGCAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGGGAAGTGGCGCAAGAGGGGCCAC 180
Qy 187 CAGATGAAGACTGTCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCCTCCCC 246
Db 181 CAGATGAAGACTGTCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCCTCCCC 240
```

```
Qy 247 CAGGCGAAGGCCCGCGAGTTTCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGCGC 306
Db 241 CAGGCGAAGGCCCGCGAGTTTCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGCGC 300
Qy 307 GAGCTGCAGGTGGCGGGGACNACCCCGCAGCGAGCGCGCGCGGAGCGGCGAGCGGAC 366
Db 301 GAGCTGCAGGTGGCGGGGACNACCCCGCAGCGAGCGCGCGCGGAGCGGCGAGCGGAC 360
Qy 367 CTGAACTTCCCGCCAGATCAGCTGTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGGGCGC 426
Db 361 CTGAACTTCCCGCCAGATCAGCTGTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGGGCGC 420
Qy 427 CAGATCAAGAGGCGCTGTGTGGACACCGCGCGCCGACGACACCTGTGTGGAGAGATGAGC 486
Db 421 CAGATCAAGAGGCGCTGTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGGGAGATGAGC 480
Qy 487 CTGCCCCGCAAGTGGAGCCCAAGATGATGGCGGGCATCGCGGCTTTCATCAAGGTGGC 546
Db 481 CTGCCCCGCAAGTGGAGCCCAAGATGATGGCGGGCATCGCGGCTTTCATCAAGGTGGC 540
Qy 547 CAGTACGACCAAGATCTGTGATCGAGATCTGGCGGCAAGAGGCCATCGGCGCTGTATC 606
Db 541 CAGTACGACCAAGATCTGTGATCGAGATCTGGCGGCAAGAGGCCATCGGCGCTGTATC 600
Qy 607 GGCCCCACCCCCGTGAAATCATCATCGGCGCCGCAACATGCTGACCCAGCTGGGCTGCACCTG 666
Db 601 GGCCCCACCCCCGTGAAATCATCATCGGCGCCGCAACATGCTGACCCAGCTGGGCTGCACCTG 660
Qy 667 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGATGAGCGGC 726
Db 661 AACTTCCCATCAGCCCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGATGAGCGGC 720
Qy 727 CCCAAGGTGAAGCAGTGGCCCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCCCTTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 780
Qy 787 GAGGAGATGGAGAGGAGGAGGCAAGATCACCAGATCGGCCCCCGAGAACCCCTACAACACC 846
Db 781 GAGGAGATGGAGAGGAGGAGGCAAGATCACCAGATCGGCCCCCGAGAACCCCTACAACACC 840
Qy 847 CCGTGTTCGCCATCAAGAAGAGGAGCAGACCAAGTGGCGGCAAGCTGGTGGACTTCGC 906
Db 841 CCGTGTTCGCCATCAAGAAGAGGAGCAGACCAAGTGGCGGCAAGCTGGTGGACTTCGC 900
Qy 907 GAGCTGAACAAGCGCACCCAGGACTTCTGCGAGGTGCGAGTGGGCATCCCCACCCCGCC 966
Db 901 GAGCTGAACAAGCGCACCCAGGACTTCTGCGAGGTGCGAGTGGGCATCCCCACCCCGCC 960
Qy 967 GGCTGAAGAAGAGAGAGCGTGCACCGTGTGACCGTGGCGGCGAGCGCTACTTTCAGCGTG 1026
Db 961 GGCTGAAGAAGAGAGAGCGTGCACCGTGTGACCGTGGCGGCGAGCGCTACTTTCAGCGTG 1020
Qy 1027 CCCTGGAAGGAGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAGAG 1086
Db 1021 CCCTGGAAGGAGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAGAG 1080
Qy 1087 ACCCCGCGCATCCGCTACCAAGTACAACGTCGCCCCAGGCGCTGGAAGGGCAGCCCCCAGC 1146
Db 1081 ACCCCGCGCATCCGCTACCAAGTACAACGTCGCCCCAGGCGCTGGAAGGGCAGCCCCCAGC 1140
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCCTTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCCTTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy 1207 GTGATCTTACAGTACATGAGACCTGTGACGTTGGGAGAGCAGCTTGGAGATCGGCGAGAC 1266
Db 1201 GTGATCTTACCA-----GGCCCCCTGTAGCTGGGAGCGACCTTGGAGATCGGCGAGAC 1254
Qy 1267 CGCGCCAGAGTCCGAGGAGCTGCGCACCTGCTGGCTGGGGCTTCCACACCCCGAC 1326
Db 1255 CGCGCCAGAGTCCGAGGAGCTGCGCACCTGCTGGCTGGGGCTTCCACACCCCGAC 1314
```



Db 601 GGGCCCGCCCGTGAACATCATCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG 660  
 Qy 667 AACTTCCCATCAGCCCCATCAGAACCGTCCCGTGAAGCTGAAGCCCGGCGATCGAAGGC 726  
 Db 661 AACTTCCCATCAGCCCCATCAGAACCGTCCCGTGAAGCTGAAGCCCGGCGATCGAAGGC 720  
 Qy 727 CCCAAGGTGAAGAGTGGCCCTCGCAGAGAGAGATCAAGAGCCCTGACCGCATCTGC 786  
 Db 721 CCCAAGGTGAAGAGTGGCCCTCGCAGAGAGAGATCAAGAGCCCTGACCGCATCTGC 780  
 Qy 787 GAGCAGATGGAAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGACCCCTCAACAC 846  
 Db 781 GAGGAGATGGAAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGACCCCTCAACAC 840  
 Qy 847 CCCGTGTTCCGCATCAAGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGACTTCGC 906  
 Db 841 CCCGTGTTCCGCATCAAGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGACTTCGC 900  
 Qy 907 GAGCTGAACAGCGACCCAGGACTTCTGGAGAGTGCAGCTGGGCGATCCGCCACCCCGCC 966  
 Db 901 GAGCTGAACAGCGACCCAGGACTTCTGGAGAGTGCAGCTGGGCGATCCGCCACCCCGCC 960  
 Qy 967 GGCCTGAAGAGAGAGAGCGTGCACCGTGTGAGCTGGGCGAGCGCTACTTCAGCGTG 1026  
 Db 961 GGCCTGAAGAGAGAGAGCGTGCACCGTGTGAGCTGGGCGAGCGCTACTTCAGCGTG 1020  
 Qy 1027 CCCCTGACAGGAGACTTCCGCAAGTACACCCCTTCCACCATCCCGCAGCATCAACAAAG 1086  
 Db 1021 CCCCTGACAGGAGACTTCCGCAAGTACACCCCTTCCACCATCCCGCAGCATCAACAAAG 1080  
 Qy 1087 ACCCCCGGCACTCCCTACAGTACAAGTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGC 1146  
 Db 1081 ACCCCCGGCACTCCCTACAGTACAAGTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGC 1140  
 Qy 1147 ATCTTCAGAGCAGCATGACCAAGATCTTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
 Db 1141 ATCTTCAGAGCAGCATGACCAAGATCTTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200  
 Qy 1207 GTGATCTACAGTACATGAGAGCTGTACGTGGGCGAGGACCTGGAGATCGGCGAGCAC 1266  
 Db 1201 GTGATCTACCA-----GGCCCCCTGTACGTGGGCGAGGACCTGGAGATCGGCGAGCAC 1254  
 Qy 1267 CGCGCCAAAGATCGAGGAGCTGCGAAGCAGCTGTGCTGGGCTTCCACACCCCGGAC 1326  
 Db 1255 CGCGCCAAAGATCGAGGAGCTGCGAAGCAGCTGTGCTGGGCTTCCACACCCCGGAC 1314  
 Qy 1327 AAGAGCACCAAGAGAGGAGCCCTTCTGTGTGAGTGGGCTACGAGCTGCACCCCGACAAG 1386  
 Db 1315 AAGAGCACCAAGAGAGGAGCCCTTCTGTGTGAGTGGGCTACGAGCTGCACCCCGACAAG 1368  
 Qy 1387 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGTGGACCGTGAACGACATCCAG 1446  
 Db 1369 TGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGTGGACCGTGAACGACATCCAG 1428  
 Qy 1447 AAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCGATCAAGGTGCGCCAG 1506  
 Db 1429 AAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCGATCAAGGTGCGCCAG 1488  
 Qy 1507 CTGTGCAAGCTGTGCGCGGCGCAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1566  
 Db 1489 CTGTGCAAGCTGTGCGCGGCGCAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1548  
 Qy 1567 GCGGAGCTGAGCTGGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGCAGCGGCTGTAC 1626  
 Db 1549 GCGGAGCTGAGCTGGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGCAGCGGCTGTAC 1608  
 Qy 1627 TAGACCCCGCAGAGGAGCTGTGGTGGCGGAGATCCAGAGAGGAGGCGCCAGCAGTGGAC 1686  
 Db 1609 TAGACCCCGCAGAGGAGCTGTGGTGGCGGAGATCCAGAGAGGAGGCGCCAGCAGTGGAC 1668  
 Qy 1687 TACAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTGAAGGAGTGGCGG 1746

Db 1669 TACAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATGCGC 1728  
 Qy 1747 ACCGCCACACCAACGAGCTGAAGAGCTGACCGAGGCGGTGCAAGAGATCGCCATGGAG 1806  
 Db 1729 ACCGCCACACCAACGAGCTGAAGAGCTGACCGAGGCGGTGCAAGAGATCGCCATGGAG 1788  
 Qy 1807 AGCATCTGATCTGGGCGAAGACCCCAAGTTCGCTGCGCCATCCAGAGAGACCTGG 1866  
 Db 1789 AGCATCTGATCTGGGCGAAGACCCCAAGTTCGCTGCGCCATCCAGAGAGACCTGG 1848  
 Qy 1867 GAGACCTGTGTGGACCGGACTACTGTCAGGCGCACCTTGGATCCCGAGTGGAGTTCGTGAAC 1926  
 Db 1849 GAGACCTGTGTGGACCGGACTACTGTCAGGCGCACCTTGGATCCCGAGTGGAGTTCGTGAAC 1908  
 Qy 1927 ACCCCCGCCCTGTGTGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCGGCGCCGAG 1986  
 Db 1909 ACCCCCGCCCTGTGTGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCGGCGCCGAG 1968  
 Qy 1987 ACCTTCTACGTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG 2046  
 Db 1969 ACCTTCTACGTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG 2028  
 Qy 2047 ACCGACCGGCGCGCGAGAGATCGTGAGCTGACCGCTGACGAGACCAACCAAGAGACCGAG 2106  
 Db 2029 ACCGACCGGCGCGCGAGAGATCGTGAGCTGACCGCTGACGAGACCAACCAAGAGACCGAG 2088  
 Qy 2107 CTGAGGCGCATTCAGCTGGCCCTGTCAGGACAGCGCGAGCGAGGTGAACATCGTGAACGAC 2166  
 Db 2089 CTGAGGCGCATTCAGCTGGCCCTGTCAGGACAGCGCGAGCGAGGTGAACATCGTGAACGAC 2148  
 Qy 2167 AGCAGTACGCTTGGGCGATCATTCAGGCGCGAGCCCGACCAAGAGCGAGGAGCTGTG 2226  
 Db 2149 AGCAGTACGCTTGGGCGATCATTCAGGCGCGAGCCCGACCAAGAGCGAGGAGCTGTG 2208  
 Qy 2227 AACAGATCATCGAGAGCTGTCAAGAGGAGAGGTGTACCTGAGCTGGTGGTGGCCGCC 2286  
 Db 2209 AACAGATCATCGAGAGCTGTCAAGAGGAGAGGTGTACCTGAGCTGGTGGTGGCCGCC 2268  
 Qy 2287 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCCGCAAG 2346  
 Db 2269 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCCGCAAG 2328  
 Qy 2347 GTGCTGTCTTGGACGCGCATCGATGGCGGCGATCGTGTATCTACAGTACATGAGACGCTG 2406  
 Db 2329 GTGCTGTCTTGGACGCGCATCGATGGCGGCGATCGTGTATCTACAGTACATGAGACGCTG 2388  
 Qy 2407 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463  
 Db 2389 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2445

RESULT 10

US-10-435-9  
 ; Sequence 9, Application US/10190435  
 ; Publication No. US20030143248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZUR MEGEDE, Jan  
 ; APPLICANT: BARNETT, Susan W.  
 ; APPLICANT: LIAN, Ying  
 ; APPLICANT: ENGELBRECHT, Susan  
 ; APPLICANT: VAN RENSBURG, Estrelita J.  
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
 ; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
 ; FILE REFERENCE: PP18133.003 / 2302-18133  
 ; CURRENT APPLICATION NUMBER: US/10190.435  
 ; NUMBER OF SEQ ID NOS: 319  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 3930  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:



OTHER INFORMATION: Description of Artificial Sequence: GagCompPolmut_C									
US-10-190-435-9									
Query Match	97.0%	Score 2394.8;	DB 15;	Length 3930;					
Best Local Similarity	99.28;	Pred. No. 0;							
Matches 2431; Conservative	0;	Mismatches	7;	Indels	12;	Gaps	2;		
QY	14	TGGCCGAGGCCATGAGCCAGGCCACCAAGCCCAACATCTCTGATGACGGCAGCAACTTCA	73						
DB	1487	TCGCCGAGGCCATGAGCCAGGCCACCAAGCCCAACATCTCTGATGACGGCAGCAACTTCA	1546						
QY	74	AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGGCCATTCGCCCGCA	133						
DB	1547	AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGGCCATTCGCCCGCA	1606						
QY	134	ACTGCGCGCCCCCGCAAGAGGGCTGCTTGGAAAGTGCAGAGAGGGGCCACCAAGATGA	193						
DB	1607	ACTGCGCGCCCCCGCAAGAGGGCTGCTTGGAAAGTGCAGAGAGGGGCCACCAAGATGA	1666						
QY	194	AGGACTGCACCGAGCGCCAGGCCCAACTTCTTCGCGAGGACCTTGGCCTTCCCCCAGGGCA	253						
DB	1667	AGGACTGCACCGAGCGCCAGGCCCAACTTCTTCGCGAGGACCTTGGCCTTCCCCCAGGGCA	1726						
QY	254	AGGCCCGGAGTTCCCGCAGCGAGCAGAAACCGCGCCAAACAGGCCCAACAGCGCGGAGTGC	313						
DB	1727	AGGCCCGGAGTTCCCGCAGCGAGCAGAAACCGCGCCAAACAGGCCCAACAGCGCGGAGTGC	1786						
QY	314	AGTGTGCGCGGCAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAGGGCACCTCTGAACCT	373						
DB	1787	AGTGTGCGCGGCAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAGGGCACCTCTGAACCT	1846						
QY	374	TCCCCCAGATCACCCCTGTGGCAGCGCCCCCTTGTGTAGCATCAAGGTGGCGGCCAGATCA	433						
DB	1847	TCCCCCAGATCACCCCTGTGGCAGCGCCCCCTTGTGTAGCATCAAGGTGGCGGCCAGATCA	1906						
QY	434	AGGAGGCCCTTGTGGACACCCGCGGCCGACGACACCGTGTCTTGGAGGAGATGAGCCTTGCCCG	493						
DB	1907	AGGAGGCCCTTGTGGACACCCGCGGCCGACGACACCGTGTCTTGGAGGAGATGAGCCTTGCCCG	1966						
QY	494	GCAAGTGGAGCCCCAGATGATCGCGGGCATCGCGCGCTTTCATCAAGGTGGCGCAGTACG	553						
DB	1967	GCAAGTGGAGCCCCAGATGATCGCGGGCATCGCGCGCTTTCATCAAGGTGGCGCAGTACG	2026						
QY	554	ACCAGATCTTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTCTCTGATCGGCCCCA	613						
DB	2027	ACCAGATCTTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTCTCTGATCGGCCCCA	2086						
QY	614	CCCCGTGAAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGTGACCTTGAACCTTCC	673						
DB	2087	CCCCGTGAAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGTGACCTTGAACCTTCC	2146						
QY	674	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGGAAGCCCGGATGGAAGGCCCAAGG	733						
DB	2147	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGGAAGCCCGGATGGAAGGCCCAAGG	2206						
QY	734	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGCGAGGAGA	793						
DB	2207	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGCGAGGAGA	2266						
QY	794	TGGAGAAGGAGGGCAAGATCAACGAATCGGCCCGCCGAGAAACCCCTTCAACAACCCCCGTGT	853						
DB	2267	TGGAGAAGGAGGGCAAGATCAACGAATCGGCCCGCCGAGAAACCCCTTCAACAACCCCCGTGT	2326						
QY	854	TCGCCATCAAGAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTTCGGCGAGCTGA	913						
DB	2327	TCGCCATCAAGAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTTCGGCGAGCTGA	2386						
QY	914	ACAAGCGCACCCAGGACTTCTTGGGAGGTGCAGCTGGGCATCCCCCAACCCCGCGGCTGA	973						
DB	2387	ACAAGCGCACCCAGGACTTCTTGGGAGGTGCAGCTGGGCATCCCCCAACCCCGCGGCTGA	2446						
QY	974	AGAAGAGAAGAGCGTACCGTGTGACCGTGGCGAGCGCCTACTTTCAGCGTGCCTCTGG	1033						

2447	AGAAAGAAGACGGTGAACCGTGCTGCA CGTGGGCACGCCCTACTTCACGGTGCCCTCGG	2500
Qy	1034 ACAGGAGATTTCCGAAGTACACC GCCTTTACCATCCCCAGCATCAAACAACGAGACCCC CGG	1093
	2507 ACGAGGACTTTCGCAAGTACACCGCCTTACCATCTCCAGCATCAAACAACGAGACCCC CGG	2566
Qy	1094 GCATCCGCTACCAAGTACAAGCTGCTGCC CCCAGGGCTGGAAGGCGAGCCCGCAGCATCTTCC	1153
Db	2567 GCATCCGCTACCAAGTACAAGCTGCTGCC CCCAGGGCTGGAAGGCGAGCCCGCAGCATCTTCC	2626
Qy	1154 AGAGCAGCATGACCAAGATCTCTGAGAC CTTCCGCGCCCGCAACCCCGAGATCGTGATCT	1213
Db	2627 AGAGCAGCATGACCAAGATCTCTGAGAC CTTCCGCGCCCGCAACCCCGAGATCGTGATCT	2686
Qy	1214 ACCAGTACATGAGCAGCATCTGTAGCTGG GCGAGCACCTTGGAGATCGGCACGACCCGCGCCA	1273
Db	2687 ACCA-----GGCCCCCTGTACGTGG CAGCGACCTGGAGATCGGCACGACCCGCGCCA	2740
Qy	1274 AGATCGAGGAGCTGCGCAAGCACCTGTG TGGCTTGGGGCTTCAACACCCCGCAACAAGAC	1333
Db	2741 AGATCGAGGAGCTGCGCAAGCACCTGTG TGGCTTGGGGCTTCAACACCCCGCAACAAGAC	2800
Qy	1334 ACCAGAAGGAGCCCCCTTCTCTGTAGTGG GTACTACGAGCTGCAACCCGCAACAGTGGACCG	1393
Db	2801 ACCAGAAGGAGCCCCCTTCTCTCTCCAT -----CGAGCTGCAACCCGCAACAGTGGACCG	2854
Qy	1394 TGCAAGCCATCGAGCTGCCGAGAGGAGAG CTGGACCGTGGAACGACATCCAGAAGCTGG	1453
Db	2855 TGCAAGCCATCGAGCTGCCGAGAGGAGAG CTGGACCGTGGAACGACATCCAGAAGCTGG	2914
Qy	1454 TGGCAGACTGAACCTGGGCGAGCCAGAN CTACCCCGGCATCAAGGTGCGGCAGCTGTGCA	1513
Db	2915 TGGCAGACTGAACCTGGGCGAGCCAGAN CTACCCCGGCATCAAGGTGCGGCAGCTGTGCA	2974
Qy	1514 AGCTGTCTGCGCGGCGCCAGGCCCTGAC GACATCTGTCGCCCTGACCGAGAGGCGCGAGC	1573
Db	2975 AGTGTCTGCGCGGCGCCAGGCCCTGAC GACATCTGTCGCCCTGACCGAGAGGCGCGAGC	3034
Qy	1574 TTGAGCTGGCGCGGAGAACCGCGAGATC CTGCGGAGCCCGTGCAACGCGGTGTACTAGACC	1633
Db	3035 TTGAGCTGGCGCGGAGAACCGCGAGATC CTGCGGAGCCCGTGCAACGCGGTGTACTAGACC	3094
Qy	1634 CCAGCAAGACCTTGTGTGGCGAGATCCAG AAGCAGGCGCCACGACGAGTGACCTTACAGA	1693
Db	3095 CCAGCAAGACCTTGTGTGGCGAGATCCAG AAGCAGGCGCCACGACGAGTGACCTTACAGA	3154
Qy	1694 TCTTACGAGGAGCCCTTCAGAACTGAGAC CCGCAAGTAGTACGCAAGATCGGCACCGCCC	1753
Db	3155 TCTTACGAGGAGCCCTTCAGAACTGAGAC CCGCAAGTAGTACGCAAGATCGGCACCGCCC	3214
Qy	1754 ACACCAACGCTGAAGACGCTGACGAGCC CGTGCAAGAGATCGGCATGGAGAGCATCG	1813
Db	3215 ACACCAACGCTGAAGACGCTGACGAGCC CGTGCAAGAGATCGGCATGGAGAGCATCG	3274
Qy	1814 TGATCTGGGCAAGACCCCAAGTTCCGCTG GCCATTCAGAAAGGAGACCTTGGGAGACCT	1873
Db	3275 TGATCTGGGCAAGACCCCAAGTTCCGCTG GCCATTCAGAAAGGAGACCTTGGGAGACCT	3334
Qy	1874 GGTGGACCGACTACTGGCAGGCCACTTGAN TCCCGAGTGGGGAGTTCTGTAAACACCCCC	1933
Db	3335 GGTGGACCGACTACTGGCAGGCCACTTGAN TCCCGAGTGGGGAGTTCTGTAAACACCCCC	3394
Qy	1934 CCTTGTGAAGCTGTGGTACAGCTGGAGAA GGAGCCCATCATCTGGCGCGCAGACCTTCT	1993
Db	3395 CCTTGTGAAGCTGTGGTACAGCTGGAGAA GGAGCCCATCATCTGGCGCGCAGACCTTCT	3454
Qy	1994 ACTGGGACGCGCGGCCAACCCGCGAGAC CAAGATTCGGCAAGCGCGGCTACGTGACCGACC	2053
Db	3455 ACTGGGACGCGCGGCCAACCCGCGAGAC CAAGATTCGGCAAGCGCGGCTACGTGACCGACC	3514
Qy	2054 GGGCGCGCGAGAGATCGTGAGCTTGNCCGAG ACCACCAACGAGAGACCGAGCTGCAGG	2113
Db	3515 GGGCGCGCGAGAGATCGTGAGCTTGNCCGAG ACCACCAACGAGAGACCGAGCTGCAGG	3574



```
Qy 2114 CCATCCAGCTGGCCCTGACGACAGCGCAGCGAGGTGAAATCGTGTACCGACAGCCAGT 2173
|
|
|
Db 3575 CCATCCAGCTGGCCCTGACGACAGCGCAGCGAGGTGAAATCGTGTACCGACAGCCAGT 3634
|
|
|
Qy 2174 ACGCCCTGGGCATCATCCAGGCCAGCCGACACAGAGCGAGCGAGCTGGTGAACAGA 2233
|
|
|
Db 3635 ACGCCCTGGGCATCATCCAGGCCAGCCGACACAGAGCGAGCGAGCTGGTGAACAGA 3694
|
|
|
Qy 2234 TCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCCGCCACAAAG 2293
|
|
|
Db 3695 TCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCCGCCACAAAG 3754
|
|
|
Qy 2294 GCATCGCGGCAACGACGAGATCGACAAAGCTGGTGTAGCAAGGGCATCCGAAGGTGTGT 2353
|
|
|
Db 3755 GCATCGCGGCAACGACGAGATCGACAAAGCTGGTGTAGCAAGGGCATCCGAAGGTGTGT 3814
|
|
|
Qy 2354 TCCTGGACGGCATCGATGGCGGCATCGTGTACTACCAAGTTCCTCGGGGCTAGCACCGGT 2463
|
|
|
Db 3815 TCCTGGACGGCATCGATGGCGGCATCGTGTACTACCAAGTTCCTCGGGGCTAGCACCGGT 3924
|
|
|
Qy 2414 GCAGCGCGGCCCTAGGATCGATTAAAGCTTCCTCGGGGCTAGCACCGGT 2463
|
|
|
Db 3875 GCAGCGCGGCCCTAGGATCGATTAAAGCTTCCTCGGGGCTAGCACCGGT 3924
|
|
|
```

## RESULT 11

```
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBERG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 10
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10
```

```
Query Match 96.9%; Score 2393.2; DB 15; Length 3930;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCGCGCCAAATCTCTGATGACGCGAGCAATTCATCA 73
|
|
|
Db 1487 TCGCCGAGGCCATGAGCCAGGCCACCGCGCCAAATCTCTGATGACGCGAGCAATTCATCA 1546
|
|
|
Qy 74 AGGCCCCCAAGCGCATATCAAGTGTCTCACTCGCGGCAAGAGGGCCACATCGCCCGCA 133
|
|
|
Db 1547 AGGCCCCCAAGCGCATATCAAGTGTCTCACTCGCGGCAAGAGGGCCACATCGCCCGCA 1696
|
|
|
Qy 134 ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGGAGGGCCACAGATGA 193
|
|
|
Db 1607 ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGGAGGGCCACAGATGA 1666
|
|
|
Qy 194 AGGACTGCACCGAGCGCCAGCCAACTTCTTTCGCGAGGACCTGGCCCTTCCCCCAGGGCA 253
|
|
|
Db 1667 AGGACTGCACCGAGCGCCAGCCAACTTCTTTCGCGAGGACCTGGCCCTTCCCCCAGGGCA 1726
|
|
|
Qy 254 AGGCCCCGAGATTCCCCCAGCGAGAGAGAAACCGCGCCAAACAGCCCCCAGCGAGCTGC 313
|
|
|
```

```
Db 1727 AGGCCCGGAGTTCCCCAGAGGAGCAGAAACCGCGCCAAACAGCCCCCAACAGCCGCGAGCTGC 1786
|
|
|
Qy 314 AGGTGCGCGGCGCAAAACCCCGCAGCGAGGCGCGCGCGCGAGCGCGCGAGCGCGCGCGCGCG 373
|
|
|
Db 1787 AGGTGCGCGGCGCAAAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846
|
|
|
Qy 374 TCCCCCAGATCACCTGTGGCAGGCGCCCTCTGTGTAGGATCAAGGTGGCGCGCGCGCGCGCGCG 433
|
|
|
Db 1847 TCCCCCAGATCACCTGTGGCAGGCGCCCTCTGTGTAGGATCAAGGTGGCGCGCGCGCGCGCG 1906
|
|
|
Qy 434 AGGAGGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
|
|
|
Db 1907 AGGAGGCGCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
|
|
|
Qy 494 GCAAGTGGAAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553
|
|
|
Db 1967 GCAAGTGGAAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026
|
|
|
Qy 554 ACAGATCTGTATCGAGATCTGCGGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613
|
|
|
Db 2027 ACCAGATCTGTATCGAGATCTGCGGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2086
|
|
|
Qy 614 CCCCCGTGAAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
|
|
|
Db 2087 CCCCCGTGAAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146
|
|
|
Qy 674 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGAAAGCCCGCGCGCGCGCGCGCGCGCGCG 733
|
|
|
Db 2147 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGCGCGCGCGCGCGCGCGCG 2206
|
|
|
Qy 734 TGAAGCAGTGGCCCTTGAACGAGGAGAGATCAAGGCGCTGTGACCGCGCATCTGCGAGGAGA 793
|
|
|
Db 2207 TGAAGCAGTGGCCCTTGAACGAGGAGAGATCAAGGCGCTGTGACCGCGCATCTGCGAGGAGA 2266
|
|
|
Qy 794 TGGAGAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853
|
|
|
Db 2267 TGGAGAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2326
|
|
|
Qy 854 TCGCCATCAAGAAGAGGAGGAGCAGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 913
|
|
|
Db 2327 TCGCCATCAAGAAGAGGAGGAGCAGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2386
|
|
|
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGACGCTGGGCGCATCCCGCGCGCGCGCGCGCGCG 973
|
|
|
Db 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGACGCTGGGCGCATCCCGCGCGCGCGCGCGCGCG 2446
|
|
|
Qy 974 AGAAGAGAGAGCGTGCACCGTGTGACGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1033
|
|
|
Db 2447 AGAAGAGAGAGCGTGCACCGTGTGACGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2506
|
|
|
Qy 1034 ACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1093
|
|
|
Db 2507 ACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2566
|
|
|
Qy 1094 GCATTCGCTACCAAGTACAAAGTGTGCGCCAGGCGCTGGAAGGCGCGCGCGCGCGCGCGCGCG 1153
|
|
|
Db 2567 GCATTCGCTACCAAGTACAAAGTGTGCGCCAGGCGCTGGAAGGCGCGCGCGCGCGCGCGCGCG 2626
|
|
|
Qy 1154 AGAGCAGCATGACCAAGATCTTGAGGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1213
|
|
|
Db 2627 AGAGCAGCATGACCAAGATCTTGAGGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2686
|
|
|
Qy 1214 ACCAGTACATGAGCAGCATCTGTAGTGGCGAGCGACCTGGAGATCGGCGAGCACCGCGCGCGCA 1273
|
|
|
Db 2687 ACCA-----GGCCCCCTGTACGTGGCGAGCGACCTGGAGATCGGCGAGCACCGCGCGCGCA 2740
|
|
|
Qy 1274 AGATCGAGGAGCTGCGCAAGCACTGTGTGCTGGGCGCTTCAACACCCCGCGCAAGAGCG 1333
|
|
|
Db 2741 AGATCGAGGAGCTGCGCAAGCACTGTGTGCTGGGCGCTTCAACACCCCGCGCAAGAGCG 2800
|
|
|
Qy 1334 ACCAGAGGAGCGCCCCCTTCTGTGGATGGGCTACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1393
|
|
|
Db 2801 ACCAGAGGAGCGCCCCCTTCTGTGCCAT-----CGAGCTGCACCGCGCGCGCGCGCGCGCGCG 2854
|
|
|
```

```
QY 1394 TGCAAGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 1453
Db 2855 TGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGG 2914
QY 1454 TGGGCAAGCTGAATGTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTGCGGCAGCTGTGCA 1513
Db 2915 TGGGCAAGCTGAATGTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTGCGGCAGCTGTGCA 2974
QY 1514 AGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTGCCCTTGACCGAGAGCGCGAGC 1573
Db 2975 AGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTGCCCTTGACCGAGAGCGCGAGC 3034
QY 1574 TGGAGCTGGCGGAGAACCGCGAGATCTGTGCGAGCCGCTGACACGGCTGTACTAGACC 1633
Db 3035 TGGAGCTGGCGGAGAACCGCGAGATCTGTGCGAGCCGCTGACACGGCTGTACTAGACC 3094
QY 1634 CCAGCAAGGACCTGTGTGGCGGAGATCCAGAAGCAGGCGCCACGACCTAGTGAGCTTACCAGA 1693
Db 3095 CCAGCAAGGACCTGTGTGGCGGAGATCCAGAAGCAGGCGCCACGACCTAGTGAGCTTACCAGA 3154
QY 1694 TCTACAGAGGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGGCACCGCCC 1753
Db 3155 TCTACAGAGGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGGCACCGCCC 3214
QY 1754 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGACAGAGATCGCCATGGAGGACATCG 1813
Db 3215 ACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGACAGAGATCGCCATGGAGGATCG 3274
QY 1814 TGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAAGGAGACCTGGGAGACCT 1873
Db 3275 TGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAAGGAGACCTGGGAGACCT 3334
QY 1874 GGTGACCCGACTACTGCGAGGCCACCTGAGATCCCGAGTGGGAGTTCTGTGAACACCCCCC 1933
Db 3335 GGTGACCCGACTACTGCGAGGCCACCTGAGATCCCGAGTGGGAGTTCTGTGAACACCCCCC 3394
QY 1934 CCCTGGTGAAGCTGTGTACAGCTGAGAGAGAGGCCCATCATCTGGCGCGCGAGACCTTCT 1993
Db 3395 CCCTGGTGAAGCTGTGTACAGCTGAGAGAGAGGCCCATCATCTGGCGCGCGAGACCTTCT 3454
QY 1994 ACCTGGAACCGCGCGCCCAACCGCGGAGACCAAGATCGGCAAGGCGGCTACGTACCGACC 2053
Db 3455 ACCTGGAACCGCGCGCCCAACCGCGGAGACCAAGATCGGCAAGGCGGCTACGTACCGACC 3514
QY 2054 GGGGCGGCGAGAGATCGTGAAGCTGACCGAGACCAACACCAAGAGACCGAGCTGCAGG 2113
Db 3515 GGGGCGGCGAGAGATCGTGAAGCTGACCGAGACCAACACCAAGAGACCGAGCTGCAGG 3574
QY 2114 CCATCCAGCTGGCCCTGCGAGACAGCGGACGAGGTGAACATCTGTGACCGACAGCCAGT 2173
Db 3575 CCATCCAGCTGGCCCTGCGAGACAGCGGACGAGGTGAACATCTGTGACCGACAGCCAGT 3634
QY 2174 ACGCCCTGGGCATCATCCAGGCCCGAGCCGACAAAGAGCGAGAGCTGGTGAACAGGA 2233
Db 3635 ACGCCCTGGGCATCATCCAGGCCCGAGCCGACAAAGAGCGAGAGCTGGTGAACAGGA 3694
QY 2234 TCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACTGAGCTGGTGGTCCCGCCCAAGG 2293
Db 3695 TCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACTGAGCTGGTGGTCCCGCCCAAGG 3754
QY 2294 GCATCCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATCCGCAAGGTGCTGT 2353
Db 3755 GCATCCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATCCGCAAGGTGCTGT 3814
QY 2354 TCCTGGAACGCGCATCGATGGCGGCATCGTGTACTTACCAAGTACATGAGCAGCTGTACGTGG 2413
Db 3815 TCCTGGAACGCGCATCGATGGCGGCATCGTGTACTTACCAAGTACATGAGCAGCTGTACGTGG 3874
QY 2414 GCAGCGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db 3875 GCAGCGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924
```

```
RESULT 12
US-10-190-435-11
; Sequence 11, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna_C
US-10-190-435-11

Query Match 96.9%; Score 2393.2; DB 15; Length 3930;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

QY 14 TGGCCGAGGCCATGAGCCAGGCCACCAGCGCCCAACATCTCTGATGCGAGCGCAGCAACTTCA 73
Db 1487 TCGCCGAGGCCATGAGCCAGGCCACCAGCGCCCAACATCTCTGATGCGAGCGCAGCAACTTCA 1546
QY 74 AGGGCCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGGAGGCCACATCGCCCGCA 133
Db 1547 AGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGGAGGCCACATCGCCCGCA 1606
QY 134 ACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAGGGCCACAGATGA 193
Db 1607 ACTGCCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAGGGCCACAGATGA 1666
QY 194 AGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGGCTTCCCGCCAGGGCA 253
Db 1667 AGGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGGCTTCCCGCCAGGGCA 1726
QY 254 AGGCCCGGAGTTCCCGAGGAGCAGAACCGCGCCCAACAGCCCCACAGCGCGGAGCTGC 313
Db 1727 AGGCCCGGAGTTCCCGAGGAGCAGAACCGCGCCCAACAGCCCCACAGCGCGGAGCTGC 1786
QY 314 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGGCACCCTGAACCT 373
Db 1787 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGGCACCCTGAACCT 1846
QY 374 TCCCCCAGATCACCTGTGCGACGCGCCCTTGGTGAGCATCAAGGTGGGCGGCGCAGATCA 433
Db 1847 TCCCCCAGATCACCTGTGCGACGCGCCCTTGGTGAGCATCAAGGTGGGCGGCGCAGATCA 1906
QY 434 AGAGGCGCTGTGTGACACCGGCGCCGACGACACCGTGTGTGAGGAGATGAGCTGCGCCG 493
Db 1907 AGAGGCGCTGTGTGACACCGGCGCCGACGACACCGTGTGTGAGGAGATGAGCTGCGCCG 1966
QY 494 GCAAGTGGGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCAACAAGGTGCGCGCAGTACG 553
Db 1967 GCAAGTGGGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCAACAAGGTGCGCGCAGTACG 2026
QY 554 ACCAGATCTCTGATCGAGATCTTCGCGCAAGAGGCCATTCGCGACCGTGTGATCGGCCCA 613
Db 2027 ACCAGATCTCTGATCGAGATCTTCGCGCAAGAGGCCATTCGCGACCGTGTGATCGGCCCA 2086
QY 614 CCCCCGTGAACATCATCGGCGCCCAACATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTC 673
Db 2087 CCCCCGTGAACATCATCGGCGCCCAACATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTC 2146
```



Query Match									
Best Local Similarity 96.98; Score 2393.2; DB 15; Length 5184;									
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;									
Qy	14	TGCCCCAGGCGCATGAGCCAGCCGACACAGCCGCAACATCCCTGATGCGAGCGCAGCAACTTCA	73						
Db	2741	TCGCCGAGGCCATGAGCCAGCCGACACAGCGCCCAACATCTCTGATGCGAGCGCAACTTCA	2800						
Qy	74	AGGGCCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGGCCACATCGCCGCA	133						
Db	2801	AGGGCCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGGCCACATCGCCGCA	2860						
Qy	134	ACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAAGTGCSCAAGGAGGGCCACAGATGA	193						
Db	2861	ACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAAGTGCSCAAGGAGGGCCACAGATGA	2920						
Qy	194	AGGACTGCAACGAGCGCAGGCCAACTTCTTCCGCGAGGACCTTGGCCCTTCCCCCAGGGCA	253						
Db	2921	AGGACTGCAACGAGCGCAGGCCAACTTCTTCCGCGAGGACCTTGGCCCTTCCCCCAGGGCA	2980						
Qy	254	AGGCGCGGAGTTCCCGAGCGAGAGAA CCGCGGCCAACAGCCCCACAGCGCGAGGTGC	313						
Db	2981	AGGCGCGGAGTTCCCGAGCGAGAGAA CCGCGGCCAACAGCCCCACAGCGCGAGGTGC	3040						
Qy	314	AGGTGCGCGGCGACACCCCGCAGCGAGCGCGCGCGCGAGCCAGGCGACCCCTGAACT	373						
Db	3041	AGGTGCGCGGCGACACCCCGCAGCGAGCGCGCGCGAGCCAGGCGACCCCTGAACT	3100						
Qy	374	TCCCCCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGCGCAGATCA	433						
Db	3101	TCCCCCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGCGCAGATCA	3160						
Qy	434	AGGAGGCCCTGTGTGACAC CCGGCGCCGACACACCGTGTGTGAGAGAGATGAGCTGCGCG	493						
Db	3161	AGGAGGCCCTGTGTGACAC CCGGCGCCGACACACCGTGTGTGAGAGAGATGAGCTGCGCG	3220						
Qy	494	GCAAGTGGAGGCCAGATGATCGGGGACATCGGGGACATCGGGGCTTCATCAGGTGCGCGAGTACG	553						
Db	3221	GCAAGTGGAGGCCAGATGATCGGGGACATCGGGGACATCGGGGCTTCATCAGAGTGGCGAGTACG	3280						
Qy	554	ACCAGATCTCTGATCGAGATCTGGGCAAGAGCCCATCGGCACCGTGTGATCGGGCCCCA	613						
Db	3281	ACCAGATCTCTGATCGAGATCTGGGCAAGAGCCCATCGGCA CCGTGTGATCGGGCCCCA	3340						
Qy	614	CCCCGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAACTTCC	673						
Db	3341	CCCCCGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGCTGACCCCTGAACTTCC	3400						
Qy	674	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG	733						
Db	3401	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGGACGGCCCCAAGG	3460						
Qy	734	TGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTTGAACCGCCATCTGCGAGAGGA	793						
Db	3461	TGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGCCCTTGAACCGCCATCTGCGAGAGGA	3520						
Qy	794	TGAGAGAGGAGGCAAGATCA CCAAGATCGGCCCCGAGAA CCCCCTA CAA CACCCCGTGT	853						
Db	3521	TGAGAGAGGAGGCAAGATCA CCAAGATCGGCCCCGAGAA CCCCCTA CAA CACCCCGTGT	3580						
Qy	854	TCGCCATCAGAGAGAGGACAGCACCAAGTGGCGCAGCTGCTGGACTTCGCGGAGCTGA	913						
Db	3581	TCGCCATCAGAGAGAGGACAGCACCAAGTGGCGCAGCTGCTGGAGTTCGCGGAGCTGA	3640						
Qy	914	ACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCACCACCCCGCGGCTGA	973						
Db	3641	ACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCCACCACCCCGCGGCTGA	3700						
Qy	974	AGAAGAAGAGCGTGTGACCGTGTGACCGTGTGAGCGCGACCGCTACTTCAGGCTGCCCTTGG	1033						
Db	3701	AGAAGAAGAGCGTGTGACCGTGTGAGCGCGACCGCTACTTCAGGCTGCCCTTGG	3760						

Qy	1034	ACGAGGACTTTCGCAAGTATACCGCCTTCCATCCCGCAGCATCAACAACGAGACCCCCG	1093						
Db	3761	ACGAGGACTTTCGCAAGTATACCGCCTTCCATCCCGCAGCATCAACAACGAGACCCCCG	3820						
Qy	1094	GCATCCGCTTACCAAGTACAACGCTGCTGCCCAAGGGCTTGAAGGGCAGCCCCAGCATTTCC	1153						
Db	3821	GCATCCGCTTACCAAGTACAACGCTGCTGCCCAAGGGCTTGAAGGGCAGCCCCAGCATTTCC	3880						
Qy	1154	AGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCGCAACCCGAGATCGTGATCT	1213						
Db	3881	AGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCGCAACCCGAGATCGTGATCT	3940						
Qy	1214	ACCAGTACATGGAACCACTGTACGTGGGAGCAGCACTGGAGATCGGCACGACCGGCCA	1273						
Db	3941	ACCA-----GGCCCCCTGTAGTGGGAGCGACCTGGAGATCGGCAGCACCGGCCA	3994						
Qy	1274	AGATCAGAGAGTTCGCAAGCACTCTCTGCGCTTGGGGCTTACCA CCCCCTGCAACAAGC	1333						
Db	3995	AGATCAGAGAGTTCGCAAGCACTCTCTGCGCTTGGGGCTTACCA CCCCCTGCAACAAGC	4054						
Qy	1334	ACCAGAGAGGCCCCCTTCTCTGTGTGATGGGCTACGAGCTGCACCCGCAACAGTGGACCG	1393						
Db	4055	ACCAGAGAGGCCCCCTTCTCTGTGTGATGGGCTACGAGCTGCACCCGCAACAGTGGACCG	4108						
Qy	1394	TGCAGCCCATCGAGCTGCGCGAAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTGG	1453						
Db	4109	TGCAGCCCATCGAGCTGCGCGAAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTGG	4168						
Qy	1454	TGGCAAGCTGAACCTGGGCGAGCAGATCTTACCCCGGCATCAAGGTGGCGCAGCTGTGCA	1513						
Db	4169	TGGCAAGCTGAACCTGGGCGAGCAGATCTTACCCCGGCATCAAGGTGGCGCAGCTGTGCA	4228						
Qy	1514	AGTGTGTCGCGCGCAAGGCCCTTACCGACATCTGTGCCCTTGACCGAGGAGCGCGAGC	1573						
Db	4229	AGTGTGTCGCGCGCAAGGCCCTTACCGACATCTGTGCCCTTGACCGAGGAGCGCGAGC	4288						
Qy	1574	TGAGCTGCGCGAAGAACCGCGAGATCCTGCGAGCCCGTGCACGGCGTGTACTAGACC	1633						
Db	4289	TGAGCTGCGCGAAGAACCGCGAGATCCTGCGAGCCCGTGCACGGCGTGTACTAGACC	4348						
Qy	1634	CCAGCAAGGACCTTGTGGCGAGATCCAGAAGCAGGCGCA CCAAGTGGACCTTACAGA	1693						
Db	4349	CCAGCAAGGACCTTGTGGCGAGATCCAGAAGCAGGCGCA CCAAGTGGACCTTACAGA	4408						
Qy	1694	TCCTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC	1753						
Db	4409	TCCTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC	4468						
Qy	1754	ACACCAACGACGTGAAGCAGCTGACCGAGCCGTGCAGAAGATCGCCATGGAGAGCATCG	1813						
Db	4469	ACACCAACGACGTGAAGCAGCTGACCGAGCCGTGCAGAAGATCGCCATGGAGAGCATCG	4528						
Qy	1814	TGATCTGGGGCAAGACCCCAAGTTCCGCTTGCCCATTCAGAAAGGAGACCTGGGAGACCT	1873						
Db	4529	TGATCTGGGGCAAGACCCCAAGTTCCGCTTGCCCATTCAGAAAGGAGACCTGGGAGACCT	4588						
Qy	1874	GGTGGACCGACTTCTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCCC	1933						
Db	4589	GGTGGACCGACTTCTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCCC	4648						
Qy	1934	CCCTGTGAAGCTGTGGTATCCAGCTGGAGAGGAGCCCATCATCGGCGCGCAGACCTTCT	1993						
Db	4649	CCCTGTGAAGCTGTGGTATCCAGCTGGAGAGGAGCCCATCATCGGCGCGCAGACCTTCT	4708						
Qy	1994	ACGTGAAGCGCGCCCAACCGCGAGACCAAGATTCGGGCAAGGCCGGCTACGTGACCGACC	2053						
Db	4709	ACGTGAAGCGCGCCCAACCGCGAGACCAAGATTCGGGCAAGGCCGGCTACGTGACCGACC	4768						
Qy	2054	GGGGCGGCGAGAGATCTGTGAGCTTACCGAGACCAACCAACAGAGAGCCGAGCTGAGG	2113						
Db	4769	GGGGCGGCGAGAGATCTGTGAGCTTACCGAGACCAACCAACAGAGAGCCGAGCTGAGG	4828						
Qy	2114	CCATCCAGCTGGCCCTGCGAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCGCAGCCAGT	2173						

```
Db 4829 CCATCCAGCTGGCCCTTGAGGACAGCGCAGCGAGGTGAACATCGTGACCCAGCCAGT 4888
Qy 2174 ACGCCCTGGGCATCATCCAGGCCAGCCGCAAGAGCGAGAGGTGGTGAACAGA 2233
Db 4889 ACGCCCTGGGCATCATCCAGGCCAGCCGCAAGAGCGAGAGGTGGTGAACAGA 4948
Qy 2234 TCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTGAGTGGTGGCCGCCACAAAG 2293
Db 4949 TCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTGAGTGGTGGCCGCCACAAAG 5008
Qy 2294 GCATCGCGGCAACAGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 2353
Db 5009 GCATCGCGGCAACAGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGT 5068
Qy 2354 TCCTGGAGCGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGAGGACCTGTACGTGG 2413
Db 5069 TCCTGGAGCGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGAGGACCTGTACGTGG 5128
Qy 2414 GCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db 5129 GCAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 5178

RESULT 14
US-10-190-305A-82
; Sequence 82, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190.305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 5184
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: TatRevNefgagCpolIna C
US-10-190-305A-82

Query Match 96.9%; Score 2393.2; DB 16; Length 5184;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;

Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCGCCCAACATCTTGATCGAGCGAGCAACTTCA 73
Db 2741 TGGCCGAGGCCATGAGCCAGGCCACCGCCCAACATCTTGATCGAGCGAGCAACTTCA 2800
Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGCAAGAGGGGCCACATCGCCCGCA 133
Db 2801 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGCAAGAGGGGCCACATCGCCCGCA 2860
Qy 134 ACTGCGCGGCCCCCGCAAGAGGGCTGTGGAAAGTGGCGGCAAGAGGGGCCACAGATGA 193
Db 2861 ACTGCGCGGCCCCCGCAAGAGGGCTGTGGAAAGTGGCGGCAAGAGGGGCCACAGATGA 2920
Qy 194 AGGACTGCAACGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCCTTCCCGCAGGGCA 253
Db 2921 AGGACTGCAACGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCCCTTCCCGCAGGGCA 2980
Qy 254 AGGCCCCCGAGTTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCCAGCGCGAGCTGC 313
Db 2981 AGGCCCCCGAGTTTCCCGAGGAGCAGAAACCGCGCAACAGCCCCCAGCGCGAGCTGC 3040
Qy 314 AGGTGCGCGCGCAACACCCCGCAGCGAGGCGCGCGCGCGAGCCGCGCACCCCTGAACT 373
```

```
Db 3041 AGGTGCGCGCGCAACACCCCGCAGCGAGGCGCGCGCGCGAGCGCGCGCACCTTGAACT 3100
Qy 374 TCCCCAGAGATCACCTGTGGCAGCGCCCTTGGTGTAGCATCAAGGTGGCGCGCCAGATCA 433
Db 3101 TCCCCAGAGATCACCTGTGGCAGCGCCCTTGGTGTAGCATCAAGGTGGCGCGCCAGATCA 3160
Qy 434 AGGAGGCCCTGTGTGGACACCGGCGCCGACGACACCGTGTGGAGAGATGAGCTTGCCCG 493
Db 3161 AGGAGGCCCTGTGTGGACACCGGCGCCGACGACACCGTGTGGAGAGATGAGCTTGCCCG 3220
Qy 494 GCAAGTGGAGGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGGTGGCGCAGTACG 553
Db 3221 GCAAGTGGAGGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGGTGGCGCAGTACG 3280
Qy 554 ACCAGATCTCTGATCGAGATCTTGGCGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 613
Db 3281 ACCAGATCTCTGATCGAGATCTTGGCGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 3340
Qy 614 CCCCCTGTGAACATCATCGGCGGCAACATGTGTGACCCAGCTGGGCTGCAACCTGAACTTTC 673
Db 3341 CCCCCTGTGAACATCATCGGCGGCAACATGTGTGACCCAGCTGGGCTGCAACCTGAACTTTC 3400
Qy 674 CCATCAGCCCCATCGAGACCGTGCCTGTGAGCTGGAAGCCCGGCATGGACGCGCCCAAGG 733
Db 3401 CCATCAGCCCCATCGAGACCGTGCCTGTGAGCTGGAAGCCCGGCATGGACGCGCCCAAGG 3460
Qy 734 TGAAGCAGTGGCCCTTGCACCGAGGAGAGATCAAGGCGCTTGACGCGCATCTTGCAGAGAGA 793
Db 3461 TGAAGCAGTGGCCCTTGCACCGAGGAGAGATCAAGGCGCTTGACGCGCATCTTGCAGAGAGA 3520
Qy 794 TGAAGAGGAGGCGCAAGATCAACCAAGATCGGCGCCGAGAACCCCTTACAACACCCCGTGT 853
Db 3521 TGAAGAGGAGGCGCAAGATCAACCAAGATCGGCGCCGAGAACCCCTTACAACACCCCGTGT 3580
Qy 854 TCGCCATCAAGAGAGAGCAGACCAAGTGGCGCAGCTGGTGGACTTTCGCGAGCTGA 913
Db 3581 TCGCCATCAAGAGAGAGCAGACCAAGTGGCGCAGCTGGTGGACTTTCGCGAGCTGA 3640
Qy 914 ACAAGCGCACCCAGGACTTCTTGGAGGTGACGTGGGCGATCCCGCCACCCCGCGCCCTGA 973
Db 3641 ACAAGCGCACCCAGGACTTCTTGGAGGTGACGTGGGCGATCCCGCCACCCCGCGCCCTGA 3700
Qy 974 AGAAGAGAGAGAGCGTGACCGTGTGACGCTGGGCGACGCTACTTACAGGTGCCCCGTGG 1033
Db 3701 AGAAGAGAGAGAGCGTGACCGTGTGACGCTGGGCGACGCTACTTACAGGTGCCCCGTGG 3760
Qy 1034 ACGAGGACTTTCGCAAGTACACCGCTTCAACATCCCGCAGCATCAACAGAGACCCCG 1093
Db 3761 ACGAGGACTTTCGCAAGTACACCGCTTCAACATCCCGCAGCATCAACAGAGACCCCG 3820
Qy 1094 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCGAGGCTGGAAGGGCAGCGCCAGCATCTTCC 1153
Db 3821 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCGAGGCTGGAAGGGCAGCGCCAGCATCTTCC 3880
Qy 1154 AGAGCAGCATGAACAAGATCTTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCT 1213
Db 3881 AGAGCAGCATGAACAAGATCTTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCT 3940
Qy 1214 ACCAGTACATGGAGCAGCTGTACGTGGGCGAGCATCTGGAGATCGGCAGCACCGCGCA 1273
Db 3941 ACCA-----GGCCCCCTGTACGTGGGCGAGCATCTGGAGATCGGCAGCACCGCGCA 3994
Qy 1274 AGATCGAGGAGCTTGGCGCAAGCAGCTGTGCTGGGCTTCCAGCCCGCCAGCAAGAGC 1333
Db 3995 AGATCGAGGAGCTTGGCGCAAGCAGCTGTGCTGGGCTTCCAGCCCGCCAGCAAGAGC 4054
Qy 1334 ACCAGAGGAGGCCCTTCTTCTGTGGATGGGCTACAGAGCTGCAACCCCGAGAGTGGACCG 1393
Db 4055 ACCAGAGGAGGCCCTTCTTCTGTGCCAT-----CGAGCTGCACCCCGAGAGTGGACCG 4108
Qy 1394 TGCAGCCCATCGAGCTGCCCGAGAGGAGAGCTGGACCGTGGACCGATCCAGAGCTGG 1453
```

```
Db 4109 TGCAGCCCATCGAGCTGCCCGAGAGGAGAGCTGGACCGGTGAACGACATCCAGAGCTGG 4168
Qy 1454 TGGGCAAGCTGAAGTGGGCGAGCCAGATCTACCCCGCATCAAGGTGGCGACCTGTGCA 1513
Db 4169 TGGGCAAGCTGAAGTGGGCGAGCCAGATCTACCCCGCATCAAGGTGGCGACCTGTGCA 4228
Qy 1514 AGCTGCTGGCGGCGCCAAAGCCCTGACCCAGATCTGTCGCCCTGACCGAGAGGCGCGAGC 1573
Db 4229 AGCTGCTGGCGGCGCCAAAGCCCTGACCCAGATCTGTCGCCCTGACCGAGAGGCGCGAGC 4288
Qy 1574 TGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGGTGACCGCGGTGTACTACGACC 1633
Db 4289 TGGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGGTGACCGCGGTGTACTACGACC 4348
Qy 1634 CAGCAAGAGCACTGTGTGGCGGAGATCCAGAAAGCAGGCGCCAGCAAGTGGACCTTACGAGA 1693
Db 4349 CCAGCAAGGACCTGGTGGCGGAGATCCAGAAAGCAGGCGCCAGCAAGTGGACCTTACGAGA 4408
Qy 1694 TCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATCGCGACCGCCC 1753
Db 4409 TCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGATCGCGACCGCCC 4468
Qy 1754 ACACCAACGACGTGAAGCAGCTGACCCGAGCCCGGTGCAAGAGATCGCCATGGAGAGCATCG 1813
Db 4469 ACACCAACGACGTGAAGCAGCTGACCCGAGCCCGGTGCAAGAGATCGCCATGGAGAGCATCG 4528
Qy 1814 TGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGGAGACCTTGGGAGACCT 1873
Db 4529 TGATCTGGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGGAGACCTTGGGAGACCT 4588
Qy 1874 GGTGGACCGACTACTGGCAGGCACTGAGTCCCGAGTGGGAGTTGCTGAACACCCCCC 1933
Db 4589 GGTGGACCGACTACTGGCAGGCACTGAGTCCCGAGTGGGAGTTGCTGAACACCCCCC 4648
Qy 1934 CCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCT 1993
Db 4649 CCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCT 4708
Qy 1994 ACCTGGACCGCGCGCCCAACCGCGAGCAAGATCGGCAAGGCGCGGTACGTACCGGACC 2053
Db 4709 ACCTGGACCGCGCGCCCAACCGCGAGCAAGATCGGCAAGGCGCGGTACGTACCGGACC 4768
Qy 2054 GGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 2113
Db 4769 GGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGG 4828
Qy 2114 CCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGGTGAACATCGTGACCGACAGCCAGT 2173
Db 4829 CCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGGTGAACATCGTGACCGACAGCCAGT 4888
Qy 2174 ACGCCCTGGGCATCATCCAGGCCAGCCCGAAGAGCGAGAGGAGCTGGTGAACGAGA 2233
Db 4889 ACGCCCTGGGCATCATCCAGGCCAGCCCGAAGAGCGAGAGGAGCTGGTGAACGAGA 4948
Qy 2234 TCATCCAGAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTCCCGCCCGCAAGG 2293
Db 4949 TCATCCAGAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGTCCCGCCCGCAAGG 5008
Qy 2294 GCATCCGCGGCAACGAGCAGATCGACAAGCTGGTGAAGCAAGGCGATCCGCAAGGTGCTGT 2353
Db 5009 GCATCCGCGGCAACGAGCAGATCGACAAGCTGGTGAAGCAAGGCGATCCGCAAGGTGCTGT 5068
Qy 2354 TCTTGAACGCGATCGATGGCGGCATCGTGATCTACAGTACATGGAGGACCTGTAGCTGG 2413
Db 5069 TCTTGAACGCGATCGATGGCGGCATCGTGATCTACCAAGTACATGGAGGACCTGTAGCTGG 5128
Qy 2414 GCAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2463
Db 5129 GCAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 5178
```

RESULT 15

US-10-190-435-13

```
; Sequence 13, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSHURGT, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13
```

```
Query Match 95.7%; Score 2362.8; DB 15; Length 3531;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 2411; Conservative 0; Mismatches 27; Indels 12; Gaps 2;

Qy 14 TGGCCGAGGCGATGAGCGGAGCCAGCGCCAAACATCTGTATGCGAGCGCAGCAACTTCA 73
Db 1088 TGGCCGAGGCGATGAGCGGAGCCAGCGCCAAACATCTGTATGCGAGAGAGCAACTTTAAA 1147
Qy 74 AGGGCCCCCAAGCGCATCATCAAGTGTTCAACTGCGGCAAGGAGGCGCACATCGCCGCA 133
Db 1148 AGGGCCCCCAAGCGCATCATCAAGTGTTCAACTGCGGCAAGGAGGCGCACATCGCCGCA 1207
Qy 134 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTCGCGCAAGAGGGCGCACAGATGA 193
Db 1208 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAAGTCGCGCAAGAGGGCGCACAGATGA 1267
Qy 194 AGGACTGCAACCGAGCGCGAGGCGCAACCTTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGCA 253
Db 1268 AGGACTGCAACCGAGCGCGAGGCGCAACCTTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGCA 1327
Qy 254 AGCGCCGCGAGTTCCCGAGCGAGCAAGCCGCGCCAAACAGCCCAACAGCGCGAGCTGC 313
Db 1328 AGCGCCGCGAGTTCCCGAGCGAGCAAGCCGCGCCAAACAGCCCAACAGCGCGAGCTGC 1387
Qy 314 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
Db 1388 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1447
Qy 374 TCCCCAGATCACCTGTGGCAGCGCCCTTGTGTAGCATCAAGTGGGCGGCGCAGATCA 433
Db 1448 TCCCCAGATCACCTGTGGCAGCGCCCTTGTGTAGCATCAAGTGGGCGGCGCAGATCA 1507
Qy 434 AGGAGCCCTGTCTGGACACCGCGCGCGAGCACCGTGTCTGGAGGAGATCAGCTGCGCCG 493
Db 1508 AGGAGCCCTGTCTGGACACCGCGCGCGAGCACCGTGTCTGGAGGAGATCAGCTGCGCCG 1567
Qy 494 GCAAGTGGAGCCCAAGATGATCGGCGGCATCGCGCGCTTTCATCAAGGTGGCGGAGTACG 553
Db 1568 GCAAGTGGAGCCCAAGATGATCGGCGGCATCGGCGCTTTCATCAAGGTGGCGGAGTACG 1627
Qy 554 ACCAGATCTGTATCGAGATCTGGCGCAAGAGGCGCATCGGACCGCTGTATCGGCCCCA 613
Db 1628 ACCAGATCTGTATCGAGATCTGGCGCAAGAGGCGCATCGGACCGCTGTATCGGCCCCA 1687
Qy 614 CCCCCGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCTGAACCTTC 673
Db 1688 CCCCCGTGAACATCATCGGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCTGAACCTTC 1747
Qy 674 CCATCAGCCCCATCGAGACCGTCCCGTGAAGTGAAGCCCGCGCATGGAGCGGCCCAAGG 733
```

Db 1748 CCATCAGCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGCATGGAACGGCCCCCAAGG 1807  
Qy 734 TGAAGAGTGGCCCTGTGACGAGAGAGATCAAGGGCCCTGACGCCATCTGCGAGGAGA 793  
Db 1808 TGAAGCAGTGGCCCTGTGACGAGAGAGATCAAGGGCCCTGACGCCATCTGCGAGGAGA 1867  
Qy 794 TGGAGAGGGGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTACAAACCCCGCTGT 853  
Db 1868 TGGAGAGGGGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTACAAACCCCGCTGT 1927  
Qy 854 TCGCCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 913  
Db 1928 TCGCCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 1987  
Qy 914 ACAAGCGACCCAGGACTTCTGGAGGTGCACTGGGCAATCCCCACCCCGCGGCTGA 973  
Db 1988 ACAAGCGACCCAGGACTTCTGGAGGTGCACTGGGCAATCCCCACCCCGCGGCTGA 2047  
Qy 974 AGAAGAGAGAGCGTGAACCGTGTGGAACGTGGGCGAGCGCTACTTCAGCGTGGCCCTGG 2107  
Db 2048 AGAAGAGAGAGCGTGAACCGTGTGGAACGTGGGCGAGCGCTACTTCAGCGTGGCCCTGG 2107  
Qy 1034 ACAGGAGCTTCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCGG 1093  
Db 2108 ACAGGAGCTTCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCGG 2167  
Qy 1094 GCATCCGCTACCACTGACCAAGTGTGCTGCCCAAGGCTGGAAGGGCAGCCCCCAGCATCTTCC 1153  
Db 2168 GCATCCGCTACCACTGACCAAGTGTGCTGCCCAAGGCTGGAAGGGCAGCCCCCAGCATCTTCC 2227  
Qy 1154 AGAGCAGATGACCAAGATCTTGAGGCTTCCCGGCCCGCAACCCCGAGATCTGTATCT 1213  
Db 2228 AGAGCAGATGACCAAGATCTTGAGGCTTCCCGGCCCGCAACCCCGAGATCTGTATCT 2287  
Qy 1214 ACCAGTACATGGAGCAGCTGTACCTGGGACGACCTGGAGATCGGCAGCACCGCGCCA 1273  
Db 2288 ACCA-----GGCCCCCTGTACTGTGGCAGCAGCTTGGAGATCGGCAGCACCGCGCCA 2341  
Qy 1274 AGATCGAGGAGCTGCGCAAGCAGCTGTGCTGGGCTTCAACACCCCGCAAGAAGC 1333  
Db 2342 AGATCGAGGAGCTGCGCAAGCAGCTGTGCTGGGCTTCAACACCCCGCAAGAAGC 2401  
Qy 1334 ACAGAGAGAGCCCCCTTCTGTGATGGGCTACGAGCTGCACCCCGCAAGTGGACCG 1393  
Db 2402 ACCAGAGAGAGCCCCCTTCTGTGCCAT-----CGAGCTGCACCCCGCAAGTGGACCG 2455  
Qy 1394 TGCAGCCCATCGAGCTGCGAGAGAGAGCTGGAACCGTGAAAGCATCCAGAGCTGG 1453  
Db 2456 TGCAGCCCATCGAGCTGCGAGAGAGAGCTGGAACCGTGAAAGCATCCAGAGCTGG 2515  
Qy 1454 TGGCAAGCTGAACCTGGGCCAGCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGCA 1513  
Db 2516 TGGCAAGCTGAACCTGGGCCAGCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGCA 2575  
Qy 1514 AGCTGTGCGGCGCGCAAGCCCTGACCGACATCTGTGCCCTTGAACCGAGAGGCCGAGC 1573  
Db 2576 AGCTGTGCGGCGCGCAAGCCCTGACCGACATCTGTGCCCTTGAACCGAGAGGCCGAGC 2635  
Qy 1574 TGGAGCTGGCGGAGAACCGCGAGATCTGCGGAGCCCGTGCACGGCGTGTACTACGACC 1633  
Db 2636 TGGAGCTGGCGGAGAACCGCGAGATCTGCGGAGCCCGTGCACGGCGTGTACTACGACC 2695  
Qy 1634 CCAGCAAGGACCTGGTGGCCGAGATCCAGAGAGAGGCGCACGACAGTGGACCTTACCAGA 1693  
Db 2696 CCAGCAAGGACCTGGTGGCCGAGATCCAGAGAGAGGCGCACGACAGTGGACCTTACCAGA 2755  
Qy 1694 TCTACCAAGGACCTTCAAGAACTGGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 1753  
Db 2756 TCTACCAAGGACCTTCAAGAACTGGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCC 2815  
Qy 1754 ACACCAAGCGGTGAAGCAGCTGACCGAGCCGTGCAAGAGATCGCCATGGAGAGCATCG 1813  
Db 2816 ACACCAAGCGGTGAAGCAGCTGACCGAGCCGTGCAAGAGATCGCCATGGAGAGCATCG 2875

Qy 1814 TGATCTGGGCAAGACCCCCCAAGTTTCGCCCTGCCCCATCCAGAAAGAGACCTTGGGAGACCT 1873  
Db 2876 TGATCTGGGCAAGACCCCCCAAGTTTCGCCCTGCCCCATCCAGAAAGAGACCTTGGGAGACCT 2935  
Qy 1874 GGTGGACCGACTACTTGGCAGGCCACCTTGGATCCCGGAGTGGAGTTCTGTGAACACCCCCC 1933  
Db 2936 GGTGGACCGACTACTTGGCAGGCCACCTTGGATCCCGGAGTGGAGTTCTGTGAACACCCCCC 2995  
Qy 1934 CCCTGTGAAGCTGTGTATCCAGCTGGAGAGAGCCCATCATCGGCGCCGAGACCTTCT 1993  
Db 2996 CCCTGTGAAGCTGTGTATCCAGCTGGAGAGAGCCCATCATCGGCGCCGAGACCTTCT 3055  
Qy 1994 ACCTGGACCGGCCCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACCTGACCCGACC 2053  
Db 3056 ACCTGGACCGGCCCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCTACCTGACCCGACC 3115  
Qy 2054 GGGCCCGGCAAGATCGTGAAGCTGACCGAGACCCACCAAGAGACCCGAGCTGCAGG 2113  
Db 3116 GGGCCCGGCAAGATCGTGAAGCTGACCGAGACCCACCAAGAGACCCGAGCTGCAGG 3175  
Qy 2114 CCATCCAGCTGGCCCTGACGACAGCGGAGAGGTGAACATCTGTGACCCAGACCCAGT 2173  
Db 3176 CCATCCAGCTGGCCCTGACGACAGCGGAGAGGTGAACATCTGTGACCCAGACCCAGT 3235  
Qy 2174 ACGCCCTGGGCAATCATCCAGGCCAGCCCGCAAGAGCGAGAGGCTGTGAAACCA 2233  
Db 3236 ACGCCCTGGGCAATCATCCAGGCCAGCCCGCAAGAGCGAGAGGCTGTGAAACCA 3295  
Qy 2234 TCATCCGAGCAGCTGATCAAGAAAGAGAGGTGTACTGTGAGCTGGGTGCCGCCCAAGG 2293  
Db 3296 TCATCCGAGCAGCTGATCAAGAAAGAGAGGTGTACTGTGAGCTGGGTGCCGCCCAAGG 3355  
Qy 2294 GCATCCGCGGCAACGAGCAGATCGAACAGCTGGTGAAGAGGGCATCCGCAAGGTGCTGT 2353  
Db 3356 GCATCCGCGGCAACGAGCAGATCGAACAGCTGGTGAAGAGGGCATCCGCAAGGTGCTGT 3415  
Qy 2354 TCCTGGAACGGCATCGATGGCGGCATCTGTATCTTACCAAGTACATGGAAGACCTGTACGTGG 2413  
Db 3416 TCCTGGAACGGCATCGATGGCGGCATCTGTATCTTACCAAGTACATGGAAGACCTGTACGTGG 3475  
Qy 2414 GCAGCGGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGT 2463  
Db 3476 GCAGCGGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGT 3525

Search completed: June 2, 2005, 04:23:32

Job time : 1424.77 secs



**This Page Blank (uspto)**

	Query Match	100.0%;	Score 2463;	DB 10;	Length 2463;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 2463;	Conservative	0;	Mismatches	0; Indels 0; Gaps 0
QY	1	GTCCAGGCCACCATGGCCGAGGCCCATGAGCCAGGCCACCGGCCCAACATCTCTATCGAG	60		
Db	1	GTCCAGGCCACCATGGCCGAGGCCCATGAGCCAGGCCACCGGCCCAACATCTCTATCGAG	60		

QY	61	CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGGAGGCG	120
Db	61	CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGGAGGCG	120
QY	121	CACATCGCCGCAACTCGCGCGCCCGCCCGCAAGAGGGTGTCTGGAAGTGCAGCAAGGAG	180
Db	121	CACATCGCCGCAACTCGCGCGCCCGCCCGCAAGAGGGTGTCTGGAAGTGCAGCAAGGAG	180
QY	181	GGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCCAATCTTCTCCGCGAGGACTGTGCC	240
Db	181	GGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCCAATCTTCTCCGCGAGGACTGTGCC	240
QY	241	TTCCCCCAGGCAAGGCGCCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCCACC	300
Db	241	TTCCCCCAGGCAAGGCGCCGAGTTCCTCCAGCGAGGAGAACCGCGCAACAGCCCCACC	300
QY	301	AGCCGCGAGTGCAGGTGCGCGCGCAAAACCCCGCAGCGAGGCCCGCGCGAGCCCGAG	360
Db	301	AGCCGCGAGTGCAGGTGCGCGCGCAAAACCCCGCAGCGAGGCCCGCGCGAGCCCGAG	360
QY	361	GGCACCTGAACTTCCCCCAGATCACCTGTGTGCGAGGCCCTGTGTGAGCATCAAGGTG	420
Db	361	GGCACCTGAACTTCCCCCAGATCACCTGTGTGCGAGGCCCTGTGTGAGCATCAAGGTG	420
QY	421	GGCGGCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCAGCACACCGTGTGGAGGAG	480
Db	421	GGCGGCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCAGCACACCGTGTGGAGGAG	480
QY	481	ATGAGCCTTGC CGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGCTTTCATCAAG	540
Db	481	ATGAGCCTTGC CGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGCTTTCATCAAG	540
QY	541	GTGCGCGAGTACACACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGCACCGTG	600
Db	541	GTGCGCGAGTACACACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGCACCGTG	600
QY	601	CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
QY	661	ACCTGAACTTCCCATATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG	720
Db	661	ACCTGAACTTCCCATATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG	720
QY	721	GACGCCCCAAGGTGAAGTGAAGTGGCCCTTACCGAGGAGAGATCAAGGCCCTGACCGCC	780
Db	721	GACGCCCCAAGGTGAAGTGAAGTGGCCCTTACCGAGGAGAGATCAAGGCCCTTGAACCGCC	780
QY	781	ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGGAGAACCCCTAC	840
Db	781	ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGGAGAACCCCTAC	840
QY	841	AACACCCCGGTGTTCGCCATCAAGAAAGGAAGGACAGCAACCAAGTGGCGCAAGCTGGTGAC	900
Db	841	AACACCCCGGTGTTCGCCATCAAGAAAGGAAGGACAGCAACCAAGTGGCGCAAGCTGGTGAC	900
QY	901	TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCCAC	960
Db	901	TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCCAC	960
QY	961	CCCGCGGCTGAAGAAAGAAAGAGCGTGAACCGTGTGAGCTGGCGAGCGCTACTTC	1020
Db	961	CCCGCGGCTGAAGAAAGAAAGAGCGTGAACCGTGTGAGCTGGCGAGCGCTACTTC	1020
QY	1021	AGCGTCCCTTGAACGAGGACTTTCGCAAGTACACCGCCCTTCAACATCCCCAGCATCAAC	1080
Db	1021	AGCGTCCCTTGAACGAGGACTTTCGCAAGTACACCGCCCTTCAACATCCCCAGCATCAAC	1080
QY	1081	AACGAGACCCCGGCTACCGCTACCAAGTACAAAGTGTGCTGCGCCAGGGCTGGAGGGCAGC	1140
Db	1081	AACGAGACCCCGGCTACCGCTACCAAGTACAAAGTGTGCTGCGCCAGGGCTGGAGGGCAGC	1140
QY	1141	CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCGCCCAACCCCC	1200

Db	1141	CCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCGCCCAACCCCC	1200
QY	1201	GAGATCGTATCTACACAGGCCCCCTGTAGCTGGGAGCGGACTGGAGATCGGCGAGCAC	1260
Db	1201	GAGATCGTATCTACACAGGCCCCCTGTAGCTGGGAGCGGACTGGAGATCGGCGAGCAC	1260
QY	1261	CGCGCCAAAGTGCAGGAGCTGCGCAAGCACCTGTCTCGCTGGGGCTTCAACACCCCGGAC	1320
Db	1261	CGCGCCAAAGTGCAGGAGCTGCGCAAGCACCTGTCTCGCTGGGGCTTCAACACCCCGGAC	1320
QY	1321	AAGAAGCACCAAGAGGAGCCCCCTTCTGTGTGATGGGTACGAGCTGCACCCCGCAAG	1380
Db	1321	AAGAAGCACCAAGAGGAGCCCCCTTCTGTGTGATGGGTACGAGCTGCACCCCGCAAG	1380
QY	1381	TGGAACGCTGAGCCATCGAGCTGCCGAGAGAGGAGCTGGACCTGTAACGATCCAG	1440
Db	1381	TGGAACGCTGAGCCATCGAGCTGCCGAGAGAGGAGCTGGACCTGTAACGATCCAG	1440
QY	1441	AAGCTGCTGGCAAGCTGAACCTGGGCGAGCAGATCTACCCGGCATCAAGGTGCAGCAG	1500
Db	1441	AAGCTGCTGGCAAGCTGAACCTGGGCGAGCAGATCTACCCGGCATCAAGGTGCAGCAG	1500
QY	1501	CTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCGTGCCCTCTGACCGAGGAG	1560
Db	1501	CTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCGTGCCCTCTGACCGAGGAG	1560
QY	1561	GCGAGCTGGAGCTGGCCGAGAACCGGAGATCTCTCGGCGAGGCCCTGCACCGCGTGTAC	1620
Db	1561	GCGAGCTGGAGCTGGCCGAGAACCGGAGATCTCTCGGCGAGGCCCTGCACCGCGTGTAC	1620
QY	1621	TAGCACCCAGCAAGGACCTGTGTGGCCGAGATCCAGAAGCAGGCGCCACACAGTGGACC	1680
Db	1621	TAGCACCCAGCAAGGACCTGTGTGGCCGAGATCCAGAAGCAGGCGCCACACAGTGGACC	1680
QY	1681	TACCCAGATCTACCGAGAGCCCTTCAAGAACCTGGAAGACCGCGCAAGTACGCAAGATGGCG	1740
Db	1681	TACCCAGATCTACCGAGAGCCCTTCAAGAACCTGGAAGACCGCGCAAGTACGCAAGATGGCG	1740
QY	1741	ACGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAGAGATCGCCATGGAG	1800
Db	1741	ACGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAGAGATCGCCATGGAG	1800
QY	1801	AGCATCGTATCTTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACTGTG	1860
Db	1801	AGCATCGTATCTTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACTGTG	1860
QY	1861	GAGACTGTGTGAACCACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC	1920
Db	1861	GAGACTGTGTGAACCACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC	1920
QY	1921	ACCCCGCCCTGGTGAAGCTGTGGTACCGAGTGGAGAGGCCCATCATCGGCGCGGAG	1980
Db	1921	ACCCCGCCCTGGTGAAGCTGTGGTACCGAGTGGAGAGGCCCATCATCATCGGCGCGGAG	1980
QY	1981	ACCTTCTAGTGAAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGCTG	2040
Db	1981	ACCTTCTAGTGAAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGCTG	2040
QY	2041	ACCGACCGGGCGCGCAGAAAGTCTGTGACCTGACCGAGACCAACCAAGAGAGCCGAG	2100
Db	2041	ACCGACCGGGCGCGCAGAAAGTCTGTGACCTGACCGAGACCAACCAAGAGAGCCGAG	2100
QY	2101	CTGCAGGCCATCCAGCTGGCCCTGCGAGCAGGGCAGCGAGTGAACATCTGTCAGCCGAC	2160
Db	2101	CTGCAGGCCATCCAGCTGGCCCTGCGAGCAGGGCAGCGAGTGAACATCTGTCAGCCGAC	2160
QY	2161	AGCCAGTACGCTTGGGCTCATCTCAGGCGCCAGCCCGCAAGAGCGAGGAGCGAGCTGGTG	2220
Db	2161	AGCCAGTACGCTTGGGCTCATCTCAGGCGCCAGCCCGCAAGAGCGAGGAGCGAGCTGGTG	2220
QY	2221	AACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAAAGGTGTACTGTAGCTGGGTGCCGCC	2280

```
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAGAGAGGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy 2281 CACAAGGCGCATCGCGGCAACAGCAGATCAACAGCTGTGTAGCAAGGGCATCCGCAAG 2340
Db 2281 CACAAGGCGCATCGCGGCAACAGCAGATCAACAGCTGTGTAGCAAGGGCATCCGCAAG 2340
Qy 2341 GTGCTGTCTTGGAGCGGCATCGATGGCGGCATCGTGTATCTACCAAGTACATGAGCAGCTG 2400
Db 2341 GTGCTGTCTTGGAGCGGCATCGATGGCGGCATCGTGTATCTACCAAGTACATGAGCAGCTG 2400
Qy 2401 TACGTGGCGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGTGAA 2460
Db 2401 TACGTGGCGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTTAGCACCGGTGAA 2460
Qy 2461 TTC 2463
Db 2461 TTC 2463

RESULT 2
US-10-435-44
; Sequence 44, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; TYPE: DNA
; LENGTH: 2457
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-435-44

Query Match 99.7%; Score 2455.4; DB 15; Length 2457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCAGCGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCGAGCGCCAAACATCCTGATGCAG 60
Db 1 GTCAGCGCCACCATGGCGGAGGCGCATGAGCCAGGCCACCGAGCGCCAAACATCCTGATGCAG 60

Qy 61 CGCAGCAACTCAAGGGCCCCAGGCGATCATCAAGTGTCTCACTCGGCGCAGGAGGCG 120
Db 61 CGCAGCAACTCAAGGGCCCCAGGCGATCATCAAGTGTCTCACTCGGCGCAGGAGGCG 120

Qy 121 CACATCGCCGCAACTGCGCGGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180
Db 121 CACATCGCCGCAACTGCGCGGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180

Qy 181 GGCCACAGATGAAGGACTGACCGAGCGCGCAGGCCAACTTCTTCCCGAGGAGCTTGCC 240
Db 181 GGCCACAGATGAAGGACTGACCGAGCGCGCAGGCCAACTTCTTCCCGAGGAGCTTGCC 240

Qy 241 TTCCCGCAGGCGCAGGCGCGGAGTTCCCGCAGCAGCAGACCGCGCCCAACAGCCCCACC 300
Db 241 TTCCCGCAGGCGCAGGCGCGGAGTTCCCGCAGCAGCAGACCGCGCCCAACAGCCCCACC 300

Qy 301 AGCGCGAGCTGCAAGTGGCGGGGCAACACCCCGCAGCGAGGCGCGCGCGCGCGCGAG 360
Db 301 AGCGCGAGCTGCAAGTGGCGGGGCAACACCCCGCAGCGAGGCGCGCGCGCGCGCGAG 360

Qy 361 GGCAACCTGAACTTCCCCCAGATCACCTGTGTGGCAGCGCCCCCTGTGTGATCAAGGTG 420
```

```
Db 361 GGCAACCTGAACTTCCCCCAGATCACCTGTGTGGCAGCGCCCCCTGTGTGATCAAGGTG 420
Qy 421 GGCGGCCAGATCAAGGAGGCGCTCTGTGACACCGCGCGCGCAGCACCGCTGTCTGGAGGAG 480
Db 421 GGCGGCCAGATCAAGGAGGCGCTCTGTGACACCGCGCGCGCAGCACCGCTGTCTGGAGGAG 480
Qy 481 ATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG 540
Db 481 ATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAG 540
Qy 541 GTGCGCCAGTACGACGAGATCTGTATCGAGATCTGCGCGCAAGAGGCGCATCGGCGCCGTG 600
Db 541 GTGCGCCAGTACGACGAGATCTGTATCGAGATCTGCGCGCAAGAGGCGCATCGGCGCCGTG 600
Qy 601 CTGATCGGCGCCCAACCGCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCGCCCAACCGCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGACCCGAGAGATCAAGGCGCTGACCGCC 720
Db 661 ACCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGCCTGACCCGAGAGATCAAGGCGCTGACCGCC 720
Qy 721 GACGCGCCCAAGTGAAGCAGTGGCGCCCTGACCCGAGAGAGATCAAGGCGCTGACCGCC 780
Db 721 GACGCGCCCAAGTGAAGCAGTGGCGCCCTGACCCGAGAGAGATCAAGGCGCTGACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGCGCCCGCGAGAACCCCTTAC 840
Db 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGCGCCCGCGAGAACCCCTTAC 840
Qy 841 AACACCCCGCTGTTTCGCCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCGCTGTTTCGCCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCCAC 960
Db 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCCAC 960
Qy 961 CCGCGCGCTGAAAGAAAGAGAGGCTGACCGCTGTGACGCTGGGCGAGCGCTTACTTTC 1020
Db 961 CCGCGCGCTGAAAGAAAGAGAGGCTGACCGCTGTGACGCTGGGCGAGCGCTTACTTTC 1020
Qy 1021 AGCGTGCCCTTGGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCCCAGATCAAC 1080
Db 1021 AGCGTGCCCTTGGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCCCAGATCAAC 1080
Qy 1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCTGCGCCAGGCGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCTGCGCCAGGCGCTGGAAGGGCAGC 1140
Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCC 1200
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCC 1200
Qy 1201 GAGATCGTGTATCTACCAAGCGCCCTCTGTAGTGGGCGAGCGACTTGGAGATCGGCGCAGCAC 1260
Db 1201 GAGATCGTGTATCTACCAAGCGCCCTCTGTAGTGGGCGAGCGACTTGGAGATCGGCGCAGCAC 1260
Qy 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTGTGTGGCTGGGGCTTTCACACCCCGAC 1320
Db 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTGTGTGGCTGGGGCTTTCACACCCCGAC 1320
Qy 1321 AAGAAGCACAGAGGAGCGCCCTTCTGTGTGATGGCTTACGAGCTGCACCCCGACAAG 1380
Db 1321 AAGAAGCACAGAGGAGCGCCCTTCTGTGTGATGGCTTACGAGCTGCACCCCGACAAG 1380
Qy 1381 TGAACCGTGCAGCGCCCTGAGCTGCGCGAGAGGAGAGCTGGACCGTGAACGATCCAG 1440
Db 1381 TGAACCGTGCAGCGCCCTGAGCTGCGCGAGAGGAGAGCTGGACCGTGAACGATCCAG 1440
Qy 1441 AAGTGTGTGGGCAAGCTGAATCGGCGCAGCGAGATCTTACCCCGCGCATCAAGGTGGCGCAG 1500
```

```
Db 1441 AAGCTGTTGGCAAGCTGAACTGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCCAG 1500
Qy 1501 CTGTGAAGTGTCTGGCGCGCCCAAGCCCTGACCGACATCTGTGCCCTGACCGAGGAG 1560
Db 1501 CTGTGAAGTGTCTGGCGCGCCCAAGCCCTGACCGACATCTGTGCCCTGACCGAGGAG 1560
Qy 1561 GCCGAGTGGAGTGGCCGAGAACCCGAGATCTCTGGCGAGCCCGTGCACGCGCTGTAC 1620
Db 1561 GCCGAGTGGAGTGGCCGAGAACCCGAGATCTCTGGCGAGCCCGTGCACGCGCTGTAC 1620
Qy 1621 TACGACCCCAAGAACCTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAGTGGACC 1680
Db 1621 TACGACCCCAAGAACCTGTGTGGCCGAGATCCAGAAGCAGGCGCCACGACCAAGTGGACC 1680
Qy 1681 TACAGATCTACAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGC 1740
Db 1681 TACAGATCTACAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGC 1740
Qy 1741 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGCGCTGCAGAAGATCGCATGGAG 1800
Db 1741 ACCGCCACACCAACGAGCTGAAGCAGCTGACCGAGCGCTGCAGAAGATCGCATGGAG 1800
Qy 1801 AGCATCGTGTCTGGGCGAAGACCCCAAGTTCGCTCTGCCATCCAGAAGAGACCTGG 1860
Db 1801 AGCATCGTGTCTGGGCGAAGACCCCAAGTTCGCTCTGCCATCCAGAAGAGACCTGG 1860
Qy 1861 GAGACCTGTGGACCGACTACTGTGGAGCCACCTGTGATCCCGAGTGGGAGTTCTGTAAC 1920
Db 1861 GAGACCTGTGGAGCCGACTACTGTGGAGCCACCTGTGATCCCGAGTGGGAGTTCTGTAAC 1920
Qy 1921 ACCCCCCCTGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGCGCGCGAG 1980
Db 1921 ACCCCCCCTGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGCGCGCGAG 1980
Qy 1981 ACCTTCTACGTGACGCGCGCGCAACCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
Db 1981 ACCTTCTACGTGACGCGCGCGCAACCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
Qy 2041 ACCGACGGGGCGGCGAGAAAGTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACGGGGCGGCGAGAAAGTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2101 CTGACGCCATCCAGCTGGCCCTGACGAGCAGCGGAGAGTGAACATCTGTCAGCGAC 2160
Db 2101 CTGACGCCATCCAGCTGGCCCTGACGAGCAGCGGAGAGTGAACATCTGTCAGCGAC 2160
Qy 2161 AGCCAGTACGCCCTGGGCGATCATCCAGGCCAGCCGACAAAGAGCGAGCGAGCTGGTG 2220
Db 2161 AGCCAGTACGCCCTGGGCGATCATCCAGGCCAGCCGACAAAGAGCGAGCGAGCTGGTG 2220
Qy 2221 AACCATGATCTGACGAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCC 2280
Db 2221 AACCATGATCTGACGAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCGCC 2280
Qy 2281 CACAAGGGCATCGCGGCAACGAGCAGATCGACAAAGCTGGTGAAGGGCATCCGCAAG 2340
Db 2281 CACAAGGGCATCGCGGCAACGAGCAGATCGACAAAGCTGGTGAAGGGCATCCGCAAG 2340
Qy 2341 GTGCTGTCTCGACCGGCGATCGATGGCGGATCGTGTATCTACAGTACATGACGACCTG 2400
Db 2341 GTGCTGTCTCGACCGGCGATCGATGGCGGATCGTGTATCTACAGTACATGACGACCTG 2400
Qy 2401 TAGTGGGCGAGCGCGCCCTAGGATGATTAAGCTTCCCGGGGTAGCACCGGT 2457
Db 2401 TAGTGGGCGAGCGCGCCCTAGGATGATTAAGCTTCCCGGGGTAGCACCGGT 2457
```

## RESULT 3

US-10-190-305A-38

; Sequence 38, Application US/10190305A

; Publication No. US20030198621A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MESEDE, Jan

```
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; FILE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190,305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt.YM_C
US-10-190-305A-38
```

```
Query Match 99.7%; Score 2455.4; DB 16; Length 2457;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCGAGCCACCATGGCCGAGCCATGAGCCAGGCCACGAGGCCCAACATCTCTGATGCAG 60
Db 1 GTCGAGCCACCATGGCCGAGCCATGAGCCAGGCCACGAGGCCCAACATCTCTGATGCAG 60
Qy 61 CGCAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGC 120
Db 61 CGCAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGC 120
Qy 121 CACATCGCGCCAACTGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCGGCAAGGAG 180
Db 121 CACATCGCGCCAACTGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCGGCAAGGAG 180
Qy 181 GGCACACAGATGAAGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCC 240
Db 181 GGCACACAGATGAAGACTGCAACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCC 240
Qy 241 TTCTCCCGAGGCAAGGCCCGCGAGTTCCTCCAGCGAGCAACCGCCCAACAGCCCCACC 300
Db 241 TTCTCCCGAGGCAAGGCCCGCGAGTTCCTCCAGCGAGCAACCGCCCAACAGCCCCACC 300
Qy 301 AGCCGCGAGTGCAGGTGCGCGCGCAACACCCCGCAGAGGCGCGCGCGCGCGCCAG 360
Db 301 AGCCGCGAGTGCAGGTGCGCGCGCAACACCCCGCAGAGGCGCGCGCGCGCGCCAG 360
Qy 361 GGCACCTTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCCCTGTGAGCATCAAGGTG 420
Db 361 GGCACCTTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCCCTGTGAGCATCAAGGTG 420
Qy 421 GCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGACGACACCGTGTGGAGGAG 480
Db 421 GCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGACGACACCGTGTGGAGGAG 480
Qy 481 ATGAGCCTTCCCGCAAGTGAAGCCAAAGATGATCGCGGGCATCGCGCGCTTTCATCAAG 540
Db 481 ATGAGCCTTCCCGCAAGTGAAGCCAAAGATGATCGCGGGCATCGCGCGCTTTCATCAAG 540
Qy 541 GTGCGCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACGTG 600
Db 541 GTGCGCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACGTG 600
Qy 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAACCTTCCCATATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCGATG 720
Db 661 ACCCTGAACCTTCCCATATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCGATG 720
Qy 721 GACGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCC 780
Db 721 GACGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCC 780
```



QY	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGCGCAAGGAGGGC	120
Db	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGCGCAAGGAGGGC	120
QY	121	CACATCGCCGCAACTGCGCGCCCCCGCAGAAAGGGCTGCTGGAAGTGC CGCAAGGAG	180
Db	121	CACATCGCCGCAACTGCGCGCCCCCGCAGAAAGGGCTGCTGGAAGTGC CGCAAGGAG	180
QY	181	GGCCACAGATGAAGACTGCAACGAGCGCGCAGGCCAACTTCTTCGCGAGGACCTGGCC	240
Db	181	GGCCACAGATGAAGACTGCAACGAGCGCGCAGGCCAACTTCTTCGCGAGGACCTGGCC	240
QY	241	TTCCCCAGGGCAAGGCCCGCGAGTTTCCCAGGAGCAGAAACCGCGCAACAGCCCCACC	300
Db	241	TTCCCCAGGGCAAGGCCCGCGAGTTTCCCAGGAGCAGAAACCGCGCAACAGCCCCACC	300
QY	301	AGCGCGAGCTGCAAGTGGCGCGCGCAACCCCCCGCAGCGAGGCCCGCGCGCCAG	360
Db	301	AGCGCGAGCTGCAAGTGGCGCGCGCAACCCCCCGCAGCGAGGCCCGCGCGCCAG	360
QY	361	GGCACCCTGAACTTCCGCCAGATCACCTGTGTGCGAGGGCCCCCTGTGTAGCATCAAGGTG	420
Db	361	GGCACCCTGAACTTCCGCCAGATCACCTGTGTGCGAGGGCCCCCTGTGTAGCATCAAGGTG	420
QY	421	GGCGGCAGATCAAGGAGGCCCTGCTGGACACCGCGCGCGACGACACCGTGTGGAGGAG	480
Db	421	GGCGGCAGATCAAGGAGGCCCTGCTGGACACCGCGCGCGCGACGACACCGTGTGGAGGAG	480
QY	481	ATGAGCTGTCGCGGCAAGTGAAGCCCAAGATGATCGGGCGCATCGCGGCTTTCATCAAG	540
Db	481	ATGAGCTGTCGCGGCAAGTGAAGCCCAAGATGATCGGGCGCATCGCGGCTTTCATCAAG	540
QY	541	GTGCGCCAGTACACACAGATCTCTGATCGAGATCTGCGGCNAGAAAGCCCATCGCACCGTG	600
Db	541	GTGCGCCAGTACACACAGATCTCTGATCGAGATCTGCGGCNAGAAAGCCCATCGCACCGTG	600
QY	601	CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
QY	661	ACCCTGAACTTCCCATCAAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720
Db	661	ACCCTGAACTTCCCATCAAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG	720
QY	721	GACGCGCCNAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780
Db	721	GACGCGCCNAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC	780
QY	781	ATCTGCGAGGAGATGAGAAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC	840
Db	781	ATCTGCGAGGAGATGAGAAAGGAGGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC	840
QY	841	AACACCCCGGTGTCGCCATCAAGAAAGGACAGACCAAGTGGCGCAAGCTGGTGGAC	900
Db	841	AACACCCCGGTGTCGCCATCAAGAAAGGACAGACCAAGTGGCGCAAGCTGGTGGAC	900
QY	901	TTCCGCGAGCTGAACAGGCCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCAC	960
Db	901	TTCCGCGAGCTGAACAGGCCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCAC	960
QY	961	CCCGCGGCTGAAGAAAGAGAGCGTGACCGTGTGGACGTGGCGACGCGCTACTTC	1020
Db	961	CCCGCGGCTGAAGAAAGAGAGCGTGACCGTGTGGACGTGGCGACGCGCTACTTC	1020
QY	1021	AGCGTGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCCAGCATCAAC	1080
Db	1021	AGCGTGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCCAGCATCAAC	1080
QY	1081	AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCTGCTGCCCCAGGGCTGGAGGGCAGC	1140
Db	1081	AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCTGCTGCCCCAGGGCTGGAGGGCAGC	1140

QY	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCGCGCCGCAACCCC	1200
Db	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCGCGCCGCAACCCC	1200
QY	1201	GAGATCGTGATCTACCA-----GGCCCCCTGTAGCTGGGCAAGCACTTGGAGATCGGC	1254
Db	1201	GAGATCGTGATCTACCAAGTACATGGAGCAGCTGTAGCTGGGCAAGCACTTGGAGATCGGC	1260
QY	1255	CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTACACACC	1314
Db	1261	CAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGGGCTTACACACC	1320
QY	1315	CCCGACAAGAAGCACCAGAAGGAGGCCCTCTCTGTGTGATGGGCTACGAGCTGCACCCC	1374
Db	1321	CCCGACAAGAAGCACCAGAAGGAGGCCCTCTCTGTGTGATGGGCTACGAGCTGCACCCC	1380
QY	1375	GACAAGTGACCGTGCAGGCCATTCGAGCTGCCCCGAGAAAGAGAGCTGGAACCGAC	1434
Db	1381	GACAAGTGACCGTGCAGGCCATTCGAGCTGCCCCGAGAAAGAGAGCTGGAACCGAC	1440
QY	1435	ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1494
Db	1441	ATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTG	1500
QY	1495	CGCCAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTGACCGACATCGTGCCCTGACC	1554
Db	1501	CGCCAGCTGTGCAAGCTGTGCGCGCGCAAGGCCCTGACCGACATCGTGCCCTGACC	1560
QY	1555	GAGGAGCCGAGCTGAGCTGGCCGAGAACCGGAGATCTGTGGCGAGCCCTGTGACGGC	1614
Db	1561	GAGGAGCCGAGCTGAGCTGGCCGAGAACCGGAGATCTGTGGCGAGCCCTGTGACGGC	1620
QY	1615	GTGTACTACGACCCCGCAGCAAGGACCTGTGTGGCGCGAGATCCAGAAGCAGGCGCACCGAC	1674
Db	1621	GTGTACTACGACCCCGCAGCAAGGACCTGTGTGGCGCGAGATCCAGAAGCAGGCGCACCGAC	1680
QY	1675	TGGACCTACAGATCTACGAGGAGCCCTTCAAGAACTCTGAAGACCCGCGCAAGTACGCCAAG	1734
Db	1681	TGGACCTACAGATCTACGAGGAGCCCTTCAAGAACTCTGAAGACCCGCGCAAGTACGCCAAG	1740
QY	1735	ATGCGCACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCCCTGCAGAGATCGCC	1794
Db	1741	ATGCGCACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCCCTGCAGAGATCGCC	1800
QY	1795	ATGGAGAGCATCGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAG	1854
Db	1801	ATGGAGAGCATCGTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAG	1860
QY	1855	ACCTGGGAGACCTGTGTGACCGCACTACTGGCAGGCCACCTGGATCCCCAGTGGGATTC	1914
Db	1861	ACCTGGGAGACCTGTGTGACCGCACTACTGGCAGGCCACCTGGATCCCCAGTGGGATTC	1920
QY	1915	GTGAACACCCCGCTTGTGTGAGCTGTGTGATCCAGCTGGAGNAGAGGCCCATCATCGGC	1974
Db	1921	GTGAACACCCCGCTTGTGTGAGCTGTGTGATCCAGCTGGAGNAGAGGCCCATCATCGGC	1980
QY	1975	GCGGAGACCTTCTACGTGGAACCGCGCGCAACCGCGAGACCAAGATCGCAAGGCGCGC	2034
Db	1981	GCGGAGACCTTCTACGTGGAACCGCGCGCGCAACCGCGAGACCAAGATCGCAAGGCGCGC	2040
QY	2035	TAGTGACCGACCGGGGCGCGCAAGAGATCGTGAGCTTGAACCGAGACCAACCAAGAG	2094
Db	2041	TAGTGACCGACCGGGGCGCGCAAGAGATCGTGAGCTTGAACCGAGACCAACCAAGAG	2100
QY	2095	ACCGAGCTGAGGCCATTCAGCTGGGCCCTGCAGGACAGCGGCGAGCGAGTGAACATCGTG	2154
Db	2101	ACCGAGCTGAGGCCATTCAGCTGGGCCCTGCAGGACAGCGGCGAGCGAGTGAACATCGTG	2160
QY	2155	ACCGACAGCAGTACGCCCTGGGCATCATCCAGGCCCGACCCGACAAAGAGCGAGAGCGAG	2214
Db	2161	ACCGACAGCAGTACGCCCTGGGCATCATCCAGGCCCGACCCGACAAAGAGCGAGAGCGAG	2220
QY	2215	CTGGTGAACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAGGTGTACTCTGAGCTGGGTG	2274



```
Db 2221 CTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTG 2280
Qy 2275 CCOCGCCACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGCATC 2334
Db 2281 CCOCGCCACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGGCATC 2340
Qy 2335 GGCNAGTGTCTTCTTGGAGCGGCATCGATCGCGGCGATCTGTATCTACCAAGTACATGAC 2394
Db 2341 CGCAAGTGTCTTCTTGGAGCGGCATCGATCGCGGCGATCTGTATCTACCAAGTACATGAC 2400
Qy 2395 GACCTGTACGTGGGACGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2454
Db 2401 GACCTGTACGTGGGACGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Qy 2455 GGTGAATTC 2463
Db 2461 GGTGAATTC 2469

RESULT 5
US-09-899-575-32
; Sequence 32, Application US/09899575
; Publication No. US20030223961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megede, Jan
; APPLICANT: Barnett, Susan W.
; APPLICANT: Egnebrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janse
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899,575
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704
; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PR975YMMN
US-09-899-575-32

Query Match 98.9%; Score 2436.2; DB 10; Length 2457;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTCGACGCCACCATGGCGGAGGCATGAGCCAGGCCACAGCGCCAAACATCTGATGCAG 60
Db 1 GTCGACGCCACCATGGCGGAGGCATGAGCCAGGCCACAGCGCCAAACATCTGATGCAG 60
Qy 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
Qy 121 CATATCGCCGCACTGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180
Db 121 CATATCGCCGCACTGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180
Qy 181 GGCACACAGATGAAGGACTGACCGAGCGCGCAGCCAACTTCTTCCGCGAGGACCTGGCC 240
Db 181 GGCACACAGATGAAGGACTGACCGAGCGCGCAGCCAACTTCTTCCGCGAGGACCTGGCC 240
Qy 241 TTCCCCCAGGGCAAGGCCCGGAGTTTCCCGAGGAGCAGAACCGCGCCAAACAGCCCCCACC 300
Db 241 TTCCCCCAGGGCAAGGCCCGGAGTTTCCCGAGGAGCAGAACCGCGCCAAACAGCCCCCACC 300
Qy 301 AGCGCGAGTGTGAGGTGCGGGGACAAACCCCGCAGGAGCGCGCGCGGCGCGGCGCAG 360
Db 301 AGCGCGAGTGTGAGGTGCGGGGACAAACCCCGCAGGAGCGCGCGCGGCGCGGCGCAG 360
```

```
Qy 361 GGCACCCCTGAATCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTG 420
Db 361 GGCACCCCTGAATCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTG 420
Qy 421 GCGGGCCAGATCAAGGAGGCGCTGTGACACCGCGCGCGACACCGCTCTCTGGAGGAG 480
Db 421 GCGGGCCAGATCAAGGAGGCGCTGTGACACCGCGCGCGACACCGCTCTCTGGAGGAG 480
Qy 481 ATGAGCCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGGCATCGCGGGCTTTCATCAAG 540
Db 481 ATGAGCCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGGCATCGCGGGCTTTCATCAAG 540
Qy 541 GTGGCCAGTACGACGAGATCTGTATCGAGATCTGCGGGCAAGAGGCGCATCGGCACCGTG 600
Db 541 GTGGCCAGTACGACGAGATCTGTATCGAGATCTGTGCGGCAAGAGGCGCATCGGCACCGTG 600
Qy 601 CTGATCGGGCCCGCCCGCGTGAACATCATCGCGCGCAACATCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGGCCCGCCCGCGTGAACATCATCGCGCGCAACATCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAATCTTCCCATTCAGCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGATG 720
Db 661 ACCCTGAATCTTCCCATTCAGCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGGATG 720
Qy 721 GACGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC 780
Db 721 GACGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC 780
Qy 781 ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGACCCCTAC 840
Db 781 ATCTGCGAGGAGTGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGCGAGACCCCTAC 840
Qy 841 AACACCCCGCTGTTCGCCATCAAGAAGAGAGACACCAAGTGGCGCCCAAGCTGGTGAC 900
Db 841 AACACCCCGCTGTTCGCCATCAAGAAGAGAGACACCAAGTGGCGCCCAAGCTGGTGAC 900
Qy 901 TTCGCGAGTGAACAAGCGCAACCCAGGACTTCTTGGAGGTGCGAGTGGGATCCCCCAG 960
Db 901 TTCGCGAGTGAACAAGCGCAACCCAGGACTTCTTGGAGGTGCGAGTGGGATCCCCCAG 960
Qy 961 CCGCGCGCTGAGAGAGAGCGTGCAGCGTGTGCGACCTGGCGGCGCGCTACTTC 1020
Db 961 CCGCGCGCTGAGAGAGAGCGTGCAGCGTGTGCGACCTGGCGGCGCGCTACTTC 1020
Qy 1021 AGCGTGCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080
Db 1021 AGCGTGCCTTGGACGAGGACTTTCGCAAGTACACCGCTTTCACCATCCCCAGCATCAAC 1080
Qy 1081 AACGAGACCCCGGCTACCGTACCAAGTACAAAGTGTCTGCTGCCCGGAGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCTACCGTACCAAGTACAAAGTGTCTGCTGCCCGGAGGCTGGAAGGGCAGC 1140
Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGGCGCTTCCGGCGCGCGCAACCCC 1200
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGGCGCTTCCGGCGCGCGCAACCCC 1200
Qy 1201 GAGATCTGATCTTACACAGGCGCCCTGTGTAGTGGGCGAGCACCTGGAGATCGGCGCAGC 1260
Db 1201 GAGATCTGATCTTACAGGCGCCCTGTGTAGTGGGCGAGCACCTGGAGATCGGCGCAGC 1260
Qy 1261 CGCGCCAAAGTTCGAGGAGTTCGCAAGCACCTGCTGCGCTGGGGCTTTCACCAACCCCCGAC 1320
Db 1261 CGCGCCAAAGTTCGAGGAGTTCGCAAGCACCTGCTGCGCTGGGGCTTTCACCAACCCCCGAC 1320
Qy 1321 AAGAAGCACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
Db 1321 AAGAAGCACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
Qy 1381 TGACCGTGCAGCGCCCTCGAGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
Db 1375 TGACCGTGCAGCGCCCTCGAGCTGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434
```

```

1441 AAGCTGTGGCAAGCTGAACCTGGCCAGCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
1435 AAGCTGTGGCAAGCTGAACCTGGCCAGCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
1501 CTGTGCAAGCTGTGCGCGCGCCAAAGGCCCTGAACGACATCTGTGCCCTTGACCGAGAG 1560
1495 CTGTGCAAGCTGTGCGCGCGCCAAAGGCCCTGAACGACATCTGTGCCCTTGACCGAGAG 1554
1561 GCCAGCTGGAGTGGCCGAGAACCGCGAGATCCTGGCGAGCCCTGCAACGGCTGTAC 1620
1555 GCCAGCTGGAGTGGCCGAGAACCGCGAGATCCTGGCGAGCCCTGCAACGGCTGTAC 1614
1621 TAGACACCCAGCAAGCACTGTGTGGCGCGAGATCCAGAAGCAGGGCCACACCAAGTGGACC 1680
1615 TAGACACCCAGCAAGCACTGTGTGGCGCGAGATCCAGNAGCAGGGCCACACCAAGTGGACC 1674
1681 TACCAGATCTACGAGAGCCCTTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCG 1740
1675 TACCAGATCTACGAGAGCCCTTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCG 1734
1741 ACCGCCACACCAACGAGTGAAGCAAGCTGACCGAGGCCGTGCAGAGATCGCCATGGAG 1800
1735 ACCGCCACACCAACGAGTGAAGCAAGCTGACCGAGGCCGTGCAGAGATCGCCATGGAG 1794
1801 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTCGCCATCCAGAAGGAGACTGTG 1860
1795 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTCGCCATCCAGAAGGAGACTGTG 1854
1861 GAGACTGTGTGACCGCACTACTGTGGCAGGCCACTGTGATCCCGAGTGGAGTTCTGTGAAC 1920
1855 GAGACTGTGTGACCGCACTACTGTGGCAGGCCACTGTGATCCCGAGTGGAGTTCTGTGAAC 1914
1921 ACCCCCCCTGTGTAAGCTGTGTGTTACAGCTGGAGGAGGCCATCATCGGCGCGAG 1980
1915 ACCCCCCCTGTGTAAGCTGTGTGTTACAGCTGGAGGAGGCCATCATCGGCGCGAG 1974
1981 ACCTTCTAGCTGAGCGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
1975 ACCTTCTAGCTGAGCGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2034
2041 ACCGACCGGGGCGCGCAGAAAGATGTTGAGCTTGAACCGAGACCAACCAAGAAAGACCGAG 2100
2035 ACCGACCGGGGCGCGCAGAAAGATGTTGAGCTTGAACCGAGACCAACCAAGAAAGACCGAG 2094
2101 CTGAGGCGCATCCAGCTGGCCCTGCAGGACAGCGGCGAGGTTGAACATCGTGACCGAC 2160
2095 CTGAGGCGCATCCAGCTGGCCCTGCAGGACAGCGGCGAGGTTGAACATCGTGACCGAC 2154
2161 AGCCAGTACGCCCTGGGCATCATCCAGGCGCCAGCCGCAAGAGCGAGCGAGCTGGTG 2220
2155 AGCCAGTACGCCCTGGGCATCATCCAGGCGCCAGCCGCAAGAGCGAGCGAGCTGGTG 2214
2221 AACCAGATCATCAGCAGCTGTATCAAGAGGAGAGGTTACTGTAGCTGGGTGCCCGCC 2280
2215 AACCAGATCATCAGCAGCTGTATCAAGAGGAGAGGTTACTGTAGCTGGGTGCCCGCC 2274
2281 CACAAAGGCATCGGCGGCAACGAGCAGATCGCAAGCTGTTGAGCAAGGGCATCCGCAAG 2340
2275 CACAAAGGCATCGGCGGCAACGAGCAGATCGCAAGCTGTTGAGCAAGGGCATCCGCAAG 2334
2341 GTGCTGTTCTCGGACGGCATCGATGGCGGATCGTGTATCTACCAAGTACATGGAACACCTG 2400
2335 GTGCTGTTCTCGGACGGCATCGATGGCGGATCGTGTATCTACCAAGTACATGGAACACCTG 2394
2401 TAGCTGGGACCGGCGCCCTTAGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAA 2460
2395 TAGCTGGGACCGGCGCCCTTAGATCGATTAAAGCTTCCCGGGGCTAGCACCGGTGAA 2454
2461 TTC 2463
2455 TTC 2457

```

```

RESULT 6
US-10-190-435-45
; Sequence 45, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 45
; LENGTH: 2457
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt_C
US-10-190-435-45

Query Match      98.7%;   Score 2430.2;   DB 15;   Length 2457;
Best Local Similarity 99.6%;   Pred. No. 0;
Matches 2448;   Conservative 0;   Mismatches 3;   Indels 6;   Gaps 1;

QY 7  GCCACCATGGCGAGGCCATGAGCCAGGCGCACGAGGCCCAACATCTGTATGAGCGGAGC 66
DB 1  GCCACCATGGCGAGGCCATGAGCCAGGCGCACGAGGCCCAACATCTGTATGAGCGGAGC 60

QY 67 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGGCGGAGAGGGCCACATC 126
DB 61 AACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGGCGGAGAGGGCCACATC 120

QY 127 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGTGTCTGGAAGTGGCGGCAAGAGGGCCAC 186
DB 121 GCCCGCAACTGCGCGCCCGCCCGCAAGAGGGTGTCTGGAAGTGGCGGCAAGAGGGCCAC 180

QY 187 CAGATGAAGGACTGTCAACGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGTGGCTTCCCC 246
DB 181 CAGATGAAGGACTGTCAACGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGTGGCTTCCCC 240

QY 247 CAGGGCAAGGGCCGCGAGTTCCCGCAGCGAGCAGAACCGCGCCCAACAGCCCCACAGCCGC 306
DB 241 CAGGGCAAGGGCCGCGAGTTCCCGCAGCGAGCAGAACCGCGCCCAACAGCCCCACAGCCGC 300

QY 307 GAGCTGACAGTGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGGGCAAC 366
DB 301 GAGCTGACAGTGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGGGCAAC 360

QY 367 CTGAATCTTCCCCAGATCAACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGGCGGC 426
DB 361 CTGAATCTTCCCCAGATCAACCTGTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTGGGCGGC 420

QY 427 CAGATCAAGAGGGCCCTGTCTGGACACCGGGCGCGACACACCGTGTGGAGGAGATGAGC 486
DB 421 CAGATCAAGAGGGCCCTGTCTGGACACCGGGCGCGACACACCGTGTGGAGGAGATGAGC 480

QY 487 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCG 546
DB 481 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCG 540

QY 547 CAGTACGACCAAGATCTGTATCGAGATCTGCGGCAAGAGGCCATCGGACACCTGTGTGATC 606
DB 541 CAGTACGACCAAGATCTGTATCGAGATCTGCGGCAAGAGGCCATCGGACACCTGTGTGATC 600

QY 607 GGGCCCAACCCCGTGAACATCATCGGCGCCCAACATGTGACCCAGCTGGGCTGCACCCCTG 666
DB 601 GGGCCCAACCCCGTGAACATCATCGGCGCCCAACATGTGACCCAGCTGGGCTGCACCCCTG 660

```

Qy 667 AACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATCGACGGC 726  
Db 661 AACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATCGACGGC 720  
Qy 727 CCCAAGGTGAAGAGTGGCCCTGACACGAGAGAGATCAAGGCGCTGACCGCCATCTGC 786  
Db 721 CCCAAGGTGAAGAGTGGCCCTGACACGAGAGAGATCAAGGCGCTGACCGCCATCTGC 780  
Qy 787 GAGGAGTGGAGAGAGAGGCGGAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAC 846  
Db 781 GAGGAGTGGAGAGAGAGGCGGAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAC 840  
Qy 847 CCCGTGTTCCGCATCAAG 906  
Db 841 CCCGTGTTCCGCATCAAG 900  
Qy 907 GAGCTGAACAG 966  
Db 901 GAGCTGAACAG 960  
Qy 967 GGCCTGAAG 1026  
Db 961 GGCCTGAAG 1020  
Qy 1027 CCCCTGACGAG 1086  
Db 1021 CCCCTGACGAG 1080  
Qy 1087 ACCCCGCGCATCCGCTACAGTACAAAGTGTGCGCCCGAGGGCTGGAAGGGGAGCCCGAGC 1146  
Db 1081 ACCCCGCGCATCCGCTACAGTACAAAGTGTGCGCCCGAGGGCTGGAAGGGGAGCCCGAGC 1140  
Qy 1147 ATCTTCCAG 1206  
Db 1141 ATCTTCCAG 1200  
Qy 1207 GTGATCTACCA-----GGCCCCCTGTACTGGGCGAGGAGAGAGAGAGAGAGAGAGAG 1260  
Db 1201 GTGATCTACCAAGTACATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Qy 1261 CGGCGCAAGATCGAGGAG 1320  
Db 1261 CGGCGCAAGATCGAGGAG 1320  
Qy 1321 AAGAAGACCAAG 1380  
Db 1321 AAGAAGACCAAG 1380  
Qy 1381 TGGACCGTGAAG 1440  
Db 1381 TGGACCGTGAAG 1440  
Qy 1441 AAGCTGTGGGAG 1500  
Db 1441 AAGCTGTGGGAG 1500  
Qy 1501 CTGTGAAGTGTCTGCGCGGCGCAAGGCGCTGACGACATCGTCCCGCTGACCGAGAGAG 1560  
Db 1501 CTGTGAAGTGTCTGCGCGGCGCAAGGCGCTGACGACATCGTCCCGCTGACCGAGAGAG 1560  
Qy 1561 GCGGAGTGAAGTGTGCGCGGCGCAAGGCGCTGACGAGATCTGCGCGAGAGAGAGAGAGAG 1620  
Db 1561 GCGGAGTGAAGTGTGCGCGGCGCAAGGCGCTGACGAGATCTGCGCGAGAGAGAGAGAGAG 1620  
Qy 1621 TACGACCCGAG 1680  
Db 1621 TACGACCCGAG 1680  
Qy 1681 TACGAGATCTACCAAG 1740  
Db 1681 TACGAGATCTACCAAG 1740  
Qy 1741 ACCGCCCCACCAACGAG 1800

Db 1741 ACCGCCCCACCAACGAG 1800  
Qy 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCCCTTGCCTTCCAGATCCAGAGAGAGAG 1860  
Db 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCCCTTGCCTTCCAGATCCAGAGAGAGAG 1860  
Qy 1861 GAGACCTGTGGAGACCGGACTACTGCGAGGCGCAACCTTGGATCCCCGAGTGGGAGATTGTTGAAC 1920  
Db 1861 GAGACCTGTGGAGACCGGACTACTGCGAGGCGCAACCTTGGATCCCCGAGTGGGAGATTGTTGAAC 1920  
Qy 1921 ACCGCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
Db 1921 ACCGCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980  
Qy 1981 ACCTTCTACGTGGAGCGGCGGCCAACCGCGAGACCAAGATCGGCAAGAGCGCGGTACGTG 2040  
Db 1981 ACCTTCTACGTGGAGCGGCGGCCAACCGCGAGACCAAGATCGGCAAGAGCGCGGTACGTG 2040  
Qy 2041 ACCGACCGGCGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGAGAGAGAG 2100  
Db 2041 ACCGACCGGCGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGAGAGAGAG 2100  
Qy 2101 CTGACGAG 2160  
Db 2101 CTGACGAG 2160  
Qy 2161 AGCCAGTACGCGCTTGGGCGCATCATCCAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
Db 2161 AGCCAGTACGCGCTTGGGCGCATCATCCAGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
Qy 2221 AACGAGATCATGAG 2280  
Db 2221 AACGAGATCATGAG 2280  
Qy 2281 CACAAGGGCATCGCGCGCAAG 2340  
Db 2281 CACAAGGGCATCGCGCGCAAG 2340  
Qy 2341 GTGCTGTCTTGGAGCGGCGCATCGATGGCGGATCGTATCTACCAAGTACATGGAGAGAGAGAG 2400  
Db 2341 GTGCTGTCTTGGAGCGGCGCATCGATGGCGGATCGTATCTACCAAGTACATGGAGAGAGAGAG 2400  
Qy 2401 TAGCTGGGAG 2457  
Db 2401 TAGCTGGGAG 2457

## RESULT 7

US-10-190-305A-39  
; Sequence 39, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MESEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190,305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 39  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt\_C  
US-10-190-305A-39

Query Match 98.7%; Score 2430.2; DB 16; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 0;

Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;											
Qy	7	GCCACCATGGCCGAGGCATGAGCCAGGCCACCAAGCCGCAACATCTCTGATGTCAGCGCAGC	66	Qy	1087	ACCCCGGGCATCCGGTACCAAGTACAAACGTGTGCCCCCAGGGCTGGAAAGGGCGACCCCGACG	1144	Qy	1087	ACCCCGGGCATCCGGTACCAAGTACAAACGTGTGCCCCCAGGGCTGGAAAGGGCGACCCCGACG	1144
Db	1	GCCACCATGGCCGAGGCCATGAGCCAGGCCACCAAGCCGCAACATCTCTGATGTCAGCGCAGC	60	Db	1081	ACCCCGGGCATCCGGTACCAACGTGTGCCCCCAGGGCTGGAAAGGGCGACCCCGACG	1140	Qy	1147	ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCGCAACCCCGAGATC	1206
Qy	67	AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTTGGCGCAAGGAGGGCCACATC	126	Db	1141	ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCGCGCAACCCCGAGATC	1200	Qy	1207	GTGATCTACCA-----GGCCCCCTGTACTGTGGGACGACCTTGGAGATCGGCCAGCAC	1260
Db	61	AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTTGGCGCAAGGAGGGCCACATC	120	Qy	1201	GTGATCTACCACTACATGAGACGACCTGTACGTGGGACGACCTTGGAGATCGGCCAGCAC	1260	Qy	1261	CGCGCCAAAGATCGAGAGCTGCGCAAGCACTGTCTGCGTGGGGCTTACACACCCCGGAC	1320
Qy	127	GCCCGCAACTGCGCGCCCCCGCAAGAGGGCTGTCTGGAAGTGGCGCAAGGAGGGCCAC	186	Db	1201	GTGATCTACCACTACATGAGACGACCTGTACGTGGGACGACCTTGGAGATCGGCCAGCAC	1260	Qy	1261	CGCGCCAAAGATCGAGAGCTGCGCAAGCACTGTCTGCGTGGGGCTTACCAACCCCGGAC	1320
Db	121	GCCCGCAACTGCGCGCCCCCGCAAGAGGGCTGTCTGGAAGTGGCGCAAGGAGGGCCAC	180	Qy	1321	AGNAGCACCAAGAGGAGCCCTTCTGTGTGGATGGGCTACGAGCTGCACCCCGCACAG	1380	Qy	1321	AGNAGCACCAAGAGGAGCCCTTCTGTGTGGATGGGCTACGAGCTGCACCCCGCACAG	1380
Qy	187	CAGATGAAGGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCC	246	Db	1321	AGNAGCACCAAGAGGAGCCCTTCTGTGTGGATGGGCTACGAGCTGCACCCCGCACAG	1380	Qy	1381	TGACCGTGCAGCCCATCGAGCTGCCGAGAAAGAGAGCTGACCGTGAACGACATCCAG	1440
Db	181	CAGATGAAGGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCC	240	Qy	1381	TGACCGTGCAGCCCATCGAGCTGCCGAGAAAGAGAGCTGACCGTGAACGACATCCAG	1440	Qy	1441	AAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGGCCAG	1500
Qy	247	CAGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCAAACAGCCCAACAGCCGC	306	Db	1381	TGACCGTGCAGCCCATCGAGCTGCCGAGAAAGAGAGCTGACCGTGAACGACATCCAG	1440	Qy	1441	AAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGGCCAG	1500
Db	241	CAGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCAAACAGCCCAACAGCCGC	300	Qy	1441	AAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGGCCAG	1500	Qy	1501	CTGTGCAAGCTCTGCGGGCGCCCAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAG	1560
Qy	307	GAGCTGCAGGTGCGGGCGCAACCCCGCAGCGAGGCGCGCGCGCAGCGCCAGGCGCACC	366	Db	1441	AAGCTGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGGCCAG	1500	Qy	1501	CTGTGCAAGCTCTGCGGGCGCCCAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAG	1560
Db	301	GAGCTGCAGGTGCGGGCGCAACCCCGCAGCGAGGCGCGCGCGCAGCGCCAGGCGCACC	360	Qy	1501	CTGTGCAAGCTCTGCGGGCGCCCAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAG	1560	Qy	1561	GCGGAGCTGAGACTGCGCGAGAAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGCGTGTAC	1620
Qy	367	CTGAATTTCCCCAGATCACCTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC	426	Db	1501	CTGTGCAAGCTCTGCGGGCGCCCAAGGCCCTGACCGACATCGTGCCCTGACCGAGGAG	1560	Db	1561	GCGGAGCTGAGACTGCGCGAGAAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGCGTGTAC	1620
Db	361	CTGAATTTCCCCAGATCACCTGTGCGAGGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC	420	Qy	1561	GCGGAGCTGAGACTGCGCGAGAAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGCGTGTAC	1620	Qy	1621	TACGACCCCGACGAAGCACTGGTGGCCGAGATCCAGAAAGCAGGGGCCACGACCATGGGACC	1680
Qy	427	CAGATCAAGGAGGCCCTGCTGGACAACCGGCGCGCAACACCGTGTGTGGAGGAGATGAGC	486	Db	1561	GCGGAGCTGAGACTGCGCGAGAAACCGCGAGATCTCTGCGCGAGCCCGTGCACGGCGTGTAC	1620	Qy	1621	TACGACCCCGACGAAGCACTGGTGGCCGAGATCCAGAAAGCAGGGGCCACGACCATGGGACC	1680
Db	421	CAGATCAAGGAGGCCCTGCTGGACAACCGGCGCGCAACACCGTGTGTGGAGGAGATGAGC	480	Qy	1621	TACGACCCCGACGAAGCACTGGTGGCCGAGATCCAGAAAGCAGGGGCCACGACCATGGGACC	1680	Qy	1681	TACGACCTTACAGAGGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGGC	1740
Qy	487	CTGCGCGCAAGTGAAGCCCAAGATGATCGGGGCGCATCGGCGGCTTCAAGGTGGCGC	546	Db	1621	TACGACCCCGACGAAGCACTGGTGGCCGAGATCCAGAAAGCAGGGGCCACGACCATGGGACC	1680	Qy	1681	TACGACCTTACAGAGGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGGC	1740
Db	481	CTGCGCGCAAGTGAAGCCCAAGATGATCGGGGCGCATCGGCGGCTTCAAGAGTGGCGC	540	Qy	1681	TACGACCTTACAGAGGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGGC	1740	Qy	1741	ACGCGCCACACCAACGAGTGAAGCAGCTGACCGAGGCGGTGACAGAGATCGCCATGGAG	1800
Qy	547	CAGTACGACAGATCTCTGTATCGAGATCTGGGCAAGAGCCATCGGCAACCGTGTGTATC	606	Db	1681	TACGACCTTACAGAGGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGGC	1740	Qy	1741	ACGCGCCACACCAACGAGTGAAGCAGCTGACCGAGGCGGTGACAGAGATCGCCATGGAG	1800
Db	541	CAGTACGACAGATCTCTGTATCGAGATCTGGGCAAGAGCCATCGGCAACCGTGTGTATC	600	Qy	1741	ACGCGCCACACCAACGAGTGAAGCAGCTGACCGAGGCGGTGACAGAGATCGCCATGGAG	1800	Qy	1801	AGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGG	1860
Qy	607	GGCCCCACCCCGTGAAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG	666	Db	1741	ACGCGCCACACCAACGAGTGAAGCAGCTGACCGAGGCGGTGACAGAGATCGCCATGGAG	1800	Qy	1801	AGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGG	1860
Db	601	GGCCCCACCCCGTGAAACATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG	660	Qy	1801	AGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGG	1860	Qy	1861	GAGACTGTGTGACCGCATCTACTGGCAGGCGCACCTGGATCCCCCGAGTGGGAGTTCGTGAC	1920
Qy	667	AACTTCCCATCAGCCCCCATGAGACCGTGC	726	Db	1801	AGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGG	1860	Qy	1861	GAGACTGTGTGACCGCATCTACTGGCAGGCGCACCTGGATCCCCCGAGTGGGAGTTCGTGAC	1920
Db	661	AACTTCCCATCAGCCCCCATGAGACCGTGC	720	Qy	1861	GAGACTGTGTGACCGCATCTACTGGCAGGCGCACCTGGATCCCCCGAGTGGGAGTTCGTGAC	1920	Qy	1921	ACCCCCCCCTCGTGAAGCTGTGTACCCAGCTGGAGAGGAGGCCCATCATCTCGCGCCGAG	1980
Qy	727	CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCATCTGC	786	Db	1861	GAGACTGTGTGACCGCATCTACTGGCAGGCGCACCTGGATCCCCCGAGTGGGAGTTCGTGAC	1920	Qy	1921	ACCCCCCCCTCGTGAAGCTGTGTACCCAGCTGGAGAGGAGGCCCATCATCTCGCGCCGAG	1980
Db	721	CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCATCTGC	780	Qy	1921	ACCCCCCCCTCGTGAAGCTGTGTACCCAGCTGGAGAGGAGGCCCATCATCTCGCGCCGAG	1980	Qy	1981	ACCTTCTACGTGGAAGCGGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGCTGACGTG	2040
Qy	787	GAGGAGATGGAGAGAGGAGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTACAACACC	846	Db	1921	ACCCCCCCCTCGTGAAGCTGTGTACCCAGCTGGAGAGGAGGCCCATCATCTCGCGCCGAG	1980	Qy	1981	ACCTTCTACGTGGAAGCGGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGCTGACGTG	2040
Db	781	GAGGAGATGGAGAGAGGAGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTACAACACC	840	Qy	1981	ACCTTCTACGTGGAAGCGGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGCTGACGTG	2040	Qy	2041	ACCGACCGGGGCGCGCAGAAAGATCTGTAGCCTTGAACCGAGACCAACCAACAGAAAGCCGAG	2100
Qy	847	CCCGTGTTCGCGCATCAAGAGAGAGGACAGCAACCAAGTGGCGCAAGCTGTGTGACTTCCGC	906	Qy	2041	ACCGACCGGGGCGCGCAGAAAGATCTGTAGCCTTGAACCGAGACCAACCAACAGAAAGCCGAG	2100	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160
Db	841	CCCGTGTTCGCGCATCAAGAGAGAGGACAGCAACCAAGTGGCGCAAGCTGTGTGACTTCCGC	900	Db	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160
Qy	907	GAGCTGAACAGCGCACCCAGGACTCTGGGAGGTGAGCTGGGCGATCCCGCACCCCGGCC	966	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160
Db	901	GAGCTGAACAGCGCACCCAGGACTCTGGGAGGTGAGCTGGGCGATCCCGCACCCCGGCC	960	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160
Qy	967	GGCCTGAAGAGAGAGAGCGGTGACCGTGTGTGAGCGTGGCGCAGCCTTACTTTCAGCGTG	1026	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160
Db	961	GGCCTGAAGAGAGAGAGCGGTGACCGTGTGTGAGCGTGGCGCAGCCTTACTTTCAGCGTG	1020	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160
Qy	1027	CCCTGGAAGAGGACTTCCGCAAGTACACCGCTTTCACCATTCGCCAGCATCAACAGAG	1086	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160
Db	1021	CCCTGGAAGAGGACTTCCGCAAGTACACCGCTTTCACCATTCGCCAGCATCAACAGAG	1080	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160	Qy	2101	CTGCAGGCCATCCAGCTGGCCCTGAGGACAGCGGCGAGGCTGAACATCTGTGACCCGAC	2160

```
Qy 2161 AGCCGATAGCCCTGGGCGATCATCAGCGCCGAGCCCGAACAAGCGAGCGAGCTGGTG 2220
Db 2161 AGCCGATAGCCCTGGGCGATCATCAGCGCCGAGCCCGAACAAGCGAGCGAGCTGGTG 2220
Qy 2221 AACAGATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTGAGCTGGGTGCCGCC 2280
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTGAGCTGGGTGCCGCC 2280
Qy 2281 CACAAGGCGATCGGGCGCAACAGAGCAGATCGAAGCTGGTGAGCAAGGGGATCGGCAAG 2340
Db 2281 CACAAGGCGATCGGGCGCAACAGAGCAGATCGAAGCTGGTGAGCAAGGGGATCGGCAAG 2340
Qy 2341 GTGCTGTTCTTGGAGCGGCGATCGATGGCGGCGATCGTGATCTACCAAGTACATGAGCAGCTG 2400
Db 2341 GTGCTGTTCTTGGAGCGGCGATCGATGGCGGCGATCGTGATCTACCAAGTACATGAGCAGCTG 2400
Qy 2401 TACGTGGCAGCGGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457
Db 2401 TACGTGGCAGCGGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 8
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Susan
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: P18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190,435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMMW_C
US-10-190-435-43

Query Match 98.4%; Score 2422.6; DB 15; Length 2445;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 7 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCGCGCAACATCCTGATGCGAGCGCAGC 66
Db 1 GCCACCATGGCCGAGGCCATGAGCCAGGCCACCGCGCAACATCCTGATGCGAGCGCAGC 60
Qy 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGGGCAAGAGGGCCACATC 126
Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGGGCAAGAGGGCCACATC 120
Qy 127 GCCCGCAACTCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 186
Db 121 GCCCGCAACTCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCAC 180
Qy 187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGAGCTTGGCTTCCCC 246
Db 181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGAGCTTGGCTTCCCC 240
Qy 247 CAGGGCAAGGCCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCCCAGCGCGC 306
Db 241 CAGGGCAAGGCCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCCCAGCGCGC 300
Qy 307 GAGCTGAGGTGCGGGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCACC 366
Db 301 GAGCTGAGGTGCGGGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCACC 360
```

```
Qy 367 CTGAATTTCCCCAGATCACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTGGCGCGC 426
Db 361 CTGAATTTCCCCAGATCACCTGTGGCAGCGCCCCCTTGGTGAGCATCAAGGTGGCGCGC 420
Qy 427 CAGATCAAGGAGGCCCTGCTGGACACCGCGCGCGACGACACCGTCTCGAGGAGATGAGC 486
Db 421 CAGATCAAGGAGGCCCTGCTGGCCACCGCGCGCGACGACACCGTCTCGAGGAGATGAGC 480
Qy 487 CTGCCCGGCAAGTGGAAAGCCCCAAGATGATCGCGCGGCGATCGCGCGCTTTCATCAAGGTGGC 546
Db 481 CTGCCCGGCAAGTGGAAAGCCCCAAGATGATCGCGCGGCGATCGCGCGCTTTCATCAAGGTGGC 540
Qy 547 CAGTACGACAGATCTTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATC 606
Db 541 CAGTACGACAGATCTTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATC 600
Qy 607 GGGCCCGACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGT 666
Db 601 GGGCCCGACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGT 660
Qy 667 AACTTCCCATCAGCCCCCATCGAGACCGTGTGCCCTGAAAGCTGAAAGCCCGGATGACGCGC 726
Db 661 AACTTCCCATCAGCCCCCATCGAGACCGTGTGCCCTGAAAGCTGAAAGCCCGGATGACGCGC 720
Qy 727 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 780
Qy 787 GAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGGAGNACCCCTACNACACC 846
Db 781 GAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGGAGNACCCCTACNACACC 840
Qy 847 CCCGTGTTCCGCATCAAGAAGAAGGACAGCACCAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db 841 CCCGTGTTCCGCATCAAGAAGAAGGACAGCACCAGTGGCGCAAGCTGGTGGACTTCCGC 900
Qy 907 GAGCTGAACAAGCGCAACCCAGGACTTCTGGAGAGTGAGTGGGATCCCCCAACCCCGCC 966
Db 901 GAGCTGAACAAGCGCAACCCAGGACTTCTGGAGAGTGAGTGGGATCCCCCAACCCCGCC 960
Qy 967 GGCCTGAAGAAGAAGAAGGAGTGCACCGTGTGACGTGGCGGAGCGCTACTTCAGCGTG 1026
Db 961 GGCCTGAAGAAGAAGAAGGAGTGCACCGTGTGACGTGGCGGAGCGCTACTTCAGCGTG 1020
Qy 1027 CCCCTGACGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACAACGAG 1086
Db 1021 CCCCTGACGAGGAGCTTCCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACAACGAG 1080
Qy 1087 ACCCCCGGCATCCGCTACCAAGTACAACTGTCTGCCCAAGGGCTGGAAGGGCAGCCCCAGC 1146
Db 1081 ACCCCCGGCATCCGCTACCAAGTACAACTGTCTGCCCAAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCCCGAGATC 1200
Qy 1207 GTGATCTTACCAAGGCCCGCTGTACGTGGGAGCGACCTGGAGATCGGCGACGACCGCGCC 1266
Db 1201 GTGATCTTACCAAGGCCCGCTGTACGTGGGAGCGACCTGGAGATCGGCGACGACCGCGCC 1260
Qy 1267 AAGATCGAGGAGTGCACAAGCACCTCTGCGCTGGGGCTTTCACCAACCCCGCAAGAAG 1326
Db 1261 AAGATCGAGGAGTGCACAAGCACCTCTGCGCTGGGGCTTTCACCAACCCCGCAAGAAG 1320
Qy 1327 CACCAAGAAGGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGCGACCCCGCAAGTGAGCC 1386
Db 1321 CACCAAGAAGGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGCGACCCCGCAAGTGAGCC 1374
Qy 1387 GTGCAGGCCCATCGAGCTGCCCGAGAAGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1446
Db 1375 GTGCAGGCCCATCGAGCTGCCCGAGAAGAGAGCTGGACCGTGAACGACATCCAGAAGCTG 1434
```



[illegible]

## RESULT 10

REF ID: A66666  
U.S. - 10-190-435-9

US-10-190-433-9  
: Sequence 9 Application IIS/10190435

Publication No. US20030143248A1

Publication No. US20030143248A1  
CENTRAL INFORMATION

; GENERAL INFORMATION:

APPLICANT: ZUR MEGEDE, Jan

APPLICANT: BARNETT, Susan W.

; APPLICANT: LIAN, Ying

**APPLICANT: ENGELBRECHT, Susan**

APPLICANT: VAN RENSBURG, Estrelita J.

;  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE

: TITLE OF INVENTION: POLYPEPTIDES, POLY-

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

FILE REFERENCE: PFI8133.003 / 23  
CURRENT APPLICATION NUMBER: IIS/11

; CURRENT FILING DATE: 200

;; CURRENT FILING DATE: 2002-  
: NUMBER OF SEC ID NOS: 318

; NUMBER OF S

```

; SOFTWARE: Pat
; CNO ID NO 0

```

; SEQ ID NO 9

; LENGTH: 3930

```

; TYPE: DNE

```

; ORGANISM: Artificial Sequence

**; FEATURE:**

Query Match

Query Match

Best Local Similarity 99.6%; Pred. No. 0;

14 TGGCCGAGGGCCATGAGCCAGGCCACAGGCCAACATCCTGATGCAGCGCAGCAACTTCA 73



Db	1487	TCGCCGAGGCCATGAGCCAGGCCACCAAGCCCAACATCTCTGATGTCAGGCGAGCAACTTCA	1546
Qy	74	AGGGCCCCAAGCCGATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATCGCCGCA	133
Db	1547	AGGGCCCCAAGCCGATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATCGCCGCA	1606
Qy	134	ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGTGGCAAGGAGGGGCCACAGATGA	193
Db	1607	ACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGTGGCAAGGAGGGGCCACAGATGA	1666
Qy	194	AGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCCCAGGGCA	253
Db	1667	AGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCCCAGGGCA	1726
Qy	254	AGGCCCGCGATTCCCCCAGCGAGCAGAACCGCGCCCAACAGCCCAACGCGCGAGCTGC	313
Db	1727	AGGCCCGCGATTCCCCCAGCGAGCAGAACCGCGCCCAACAGCCCAACGCGCGAGCTGC	1786
Qy	314	AGGTGCGCGCGCAAAACCCCGCAGCGAGGCCGCGCGCCGAGCCGAGGCCACCTGAACT	373
Db	1787	AGGTGCGCGCGCAAAACCCCGCAGCGAGGCCGCGCGCCGAGGCCACCTGAACT	1846
Qy	374	TCCCCCAGATACCTGTGTGCAAGCGCCCTGTGTGAGCATCAAGGTGGCGGCCAGATCA	433
Db	1847	TCCCCCAGATACCTGTGTGCAAGCGCCCTGTGTGAGCATCAAGGTGGCGGCCAGATCA	1906
Qy	434	AGAGGCCCTGTGGACACCGGCGCGCAGACACCGTGTGGAGGAGATGAGCTCGCCG	493
Db	1907	AGAGGCCCTGTGGACACCGGCGCGCAGACACCGTGTGGAGGAGATGAGCTCGCCG	1966
Qy	494	GCAAGTGGAGGCCCAAGATGATCGGGGATCGCGGCTTCAACAAGTGTGCGCATACG	553
Db	1967	GCAAGTGGAGGCCCAAGATGATCGGGGATCGCGGCTTCAACAAGTGTGCGCATACG	2026
Qy	554	ACCAGATCTCTGATCGAGATCTGGGCAAGAGGCCCATCGGCACCGTGTGATCGGCCCA	613
Db	2027	ACCAGATCTCTGATCGAGATCTGGGCAAGAGGCCCATCGGCACCGTGTGATCGGCCCA	2086
Qy	614	CCCCGTGAACATATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACTTCACTTCC	673
Db	2087	CCCCGTGAACATATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACTTCACTTCC	2146
Qy	674	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGG	733
Db	2147	CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGG	2206
Qy	734	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAG	793
Db	2207	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAG	2266
Qy	794	TGAGAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGACCCCTACACACCCCGTGT	853
Db	2267	TGAGAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGACCCCTACACACCCCGTGT	2326
Qy	854	TCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGA	913
Db	2327	TCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGA	2386
Qy	914	ACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCCCACCCCGCGGCTGA	973
Db	2387	ACAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCAATCCCCACCCCGCGGCTGA	2446
Qy	974	AGAAAGAGAGCGTGACCGTCTGGAAGTGGGGCAGCGCTACTTCAAGGTGCCCCCTGG	1033
Db	2447	AGAAAGAGAGCGTGACCGTCTGGAAGTGGGGCAGCGCTACTTCAAGGTGCCCCCTGG	2506
Qy	1034	ACCAGACTTTCGCAAGTACACCGCTTCAACATCCCGACATCAACAACAGAGACCCCG	1093
Db	2507	ACCAGACTTTCGCAAGTACACCGCTTCAACATCCCGACATCAACAACAGAGACCCCG	2566
Qy	1094	GCATCCGCTACCAAGTCAACCGTCTGCCCGAGGCTGGAGGCGAGCCCGAGCATCTTCC	1153

Db	2567	GCATCCGCTACCAAGTCAACAGTGTCTGCCCGCAGGGCTGGAAGGGCAGCCCGCAGCATCTTC	2626
Qy	1154	AGAGCAGCATGACCAAGATCTCTGGAGCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	1213
Db	2627	AGAGCAGCATGACCAAGATCTCTGGAGCCTTCGCGCCCGCAACCCCGAGATCGTGATCT	2686
Qy	1214	ACCAGGCCCCCTGTACTGCTGGGCGAGCAGCTTGGAGATCGGCCAGCACCGCGCAAGATCG	1273
Db	2687	ACCAGGCCCCCTGTACTGCTGGGCGAGCAGCTTGGAGATCGGCCAGCACCGCGCAAGATCG	2746
Qy	1274	AGAGCTGGCGAAGCACTTGTCTGGGCTTACCAACCCCGCAAGAGCAGACAGCA	1333
Db	2747	AGAGCTGGCGAAGCACTTGTCTGGGCTTACCAACCCCGCAAGAGCAGACAGCA	2806
Qy	1334	AGAGGCCCCCTTCTGTGTGATGGGCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGC	1393
Db	2807	AGAGGCCCCCTTCTGTGCCAT-----CGAGCTGCACCCCGCAAGTGGACCGTGCAGC	2860
Qy	1394	CCATCGAGCTGCCCGAAGAGGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGCA	1453
Db	2861	CCATCGAGCTGCCCGAAGAGGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGCA	2920
Qy	1454	AGCTGAATGGGCGAGCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGCTGC	1513
Db	2921	AGCTGAATGGGCGAGCAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGCTGC	2980
Qy	1514	TGCGCGCGCCCAAGGCCCTGACCGACATCTGTCCCTGACCGAGGAGGCGAGCTGAGC	1573
Db	2981	TGCGCGCGCCCAAGGCCCTGACCGACATCTGTCCCTGACCGAGGAGGCGAGCTGAGC	3040
Qy	1574	TGCGCCGAGAACCGCGAGATCTCTGCGGAGCCGTGCAACCGCGTGTACTACGACCCAGCA	1633
Db	3041	TGCGCCGAGAACCGCGAGATCTCTGCGGAGCCGTGCAACCGCGTGTACTACGACCCAGCA	3100
Qy	1634	AGACCTGTGTGGCGAGATCCAGAGCAGGGCCACGACAGTGGGACCTTACAGATCTACC	1693
Db	3101	AGACCTGTGTGGCGAGATCCAGAGCAGGGCCACGACAGTGGGACCTTACAGATCTACC	3160
Qy	1694	AGGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCA	1753
Db	3161	AGGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCA	3220
Qy	1754	ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAGAGCATCGTGATCT	1813
Db	3221	ACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAGAGCATCGTGATCT	3280
Qy	1814	GGGGCAAGACCCCAAGTTCGCTCCCATCCAGAGAGAGACCTGGGAGACCTGTGGA	1873
Db	3281	GGGGCAAGACCCCAAGTTCGCTCCCATCCAGAGAGAGACCTGGGAGACCTGTGGA	3340
Qy	1874	CGGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCCCTGG	1933
Db	3341	CGGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCCCTGG	3400
Qy	1934	TGAAGCTGTGTACCAAGCTGGAGAGGCCCATCATTCGCGCGCGAGACCTTCTACGTGG	1993
Db	3401	TGAAGCTGTGTGTACCAAGCTGGAGAGGCCCATCATTCGCGCGCGAGACCTTCTACGTGG	3460
Qy	1994	ACGGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCCGGCTACGTGACCGACCGGGCC	2053
Db	3461	ACGGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCCGGCTACGTGACCGACCGGGCC	3520
Qy	2054	GGCAGAGATCTGTAGCCCTGACCGAGACCAACCAACAGAAAGCCGAGCTGCAGGCCATCC	2113
Db	3521	GGCAGAGATCTGTAGCCCTGACCGAGACCAACCAACAGAAAGCCGAGCTGCAGGCCATCC	3580
Qy	2114	AGCTGGCCCTTGCAAGCAGCGGCGAGGCTGAAACATCGTGAACCGACAGCAGTACGCC	2173
Db	3581	AGCTGGCCCTTGCAAGCAGCGGCGAGGCTGAAACATCGTGAACCGACAGCAGTACGCC	3640
Qy	2174	TGGGCAATCATCGAGGCCCGCAGCAAGAGCGAGCGAGCTGTGTAACAGCATCTCG	2233
Db	3641	TGGGCAATCATCGAGGCCCGCAGCAAGAGCGAGCGAGCTGTGTAACAGCATCTCG	3700

```
Qy 2234 AGCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGTGTCCCGCCACCAAGGGCATCG 2293
Db 3701 AGCAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGTGTCCCGCCACCAAGGGCATCG 3760
Qy 2294 GCGGCAACGAGCAGATCGACAGCTGGTGACRAGGGGCAATCCGCAAGGTGCTGTCTCGG 2353
Db 3761 GCGGCAACGAGCAGATCGACAGCTGGTGACRAGGGGCAATCCGCAAGGTGCTGTCTCGG 3820
Qy 2354 ACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGAGACCTGTACGTGGGCAGCG 2413
Db 3821 ACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGAGACCTGTACGTGGGCAGCG 3880
Qy 2414 GCGGCCCTAGATCGATTTAAAGCTTCCCGGGGTAGCACCGGT 2457
Db 3881 GCGGCCCTAGATCGATTTAAAGCTTCCCGGGGTAGCACCGGT 3924

RESULT 11
US-10-190-435-10
; Sequence 10, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBURG, Estrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PFI8133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutAtt_C
US-10-190-435-10

Query Match 98.0%; Score 2414; DB 15; Length 3930;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

Qy 14 TGGCCGAGGCATGAGCCAGCCAGCGCCACCAATCCTGATCGAGCGGCAACTTCA 73
Db 1487 TCGCCGAGGCATGAGCCAGCCAGCGCCACCAATCCTGATCGAGCGGCAACTTCA 1546
Qy 74 AGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCGAAGGCGGCAATCGCCGCA 133
Db 1547 AGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCGAAGGCGGCAATCGCCGCA 1606
Qy 134 ACTGCCCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCAGAGGGGCGCACAGATGA 193
Db 1607 ACTGCCCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCAGAGGGGCGCACAGATGA 1666
Qy 194 AGGACTGACCGAGCGCGCAGGCAACTTCTTCCGCGAGGACCTGGCCCTCCCGAGGGCA 253
Db 1667 AGGACTGACCGAGCGCGCAGGCAACTTCTTCCGCGAGGACCTGGCCCTCCCGAGGGCA 1726
Qy 254 AGGCCCCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCCACAGCGCGAGCTGC 313
Db 1727 AGGCCCCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCCACAGCGCGAGCTGC 1786
Qy 314 AGGTGCGCGGCGCAACAACCCCGCAGCGAGCGCGCGCGCGCGCGAGGCACTTGAAC 373
Db 1787 AGGTGCGCGGCGCAACAACCCCGCAGCGAGCGCGCGCGCGCGCGAGGCACTTGAAC 1846
Qy 374 TCCCCCAGATCACCTGTGGCAGCGCCCTTGTGAGATCAAGGTGGCGGCGCAGATCA 433
Db 1846 TCCCCCAGATCACCTGTGGCAGCGCCCTTGTGAGATCAAGGTGGCGGCGCAGATCA 433
```

```
Db 1847 TCCCCCAGATCACCTGTGTGGCAGCGCCCTTGTGTGAGATCAAGGTGGCGGCGCAGATCA 1906
Qy 434 AGGAGGCGCTCTGTGGACACCGCGCGCCGACGACACCGTGTCTGGAGGAGATGAGCTTGC 493
Db 1907 AGGAGGCGCTCTGTGGACACCGCGCGCCGACGACACCGTGTCTGGAGGAGATGAGCTTGC 1966
Qy 494 GCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCGCAGTACG 553
Db 1967 GCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTTCATCAAGGTGGCGCAGTACG 2026
Qy 554 ACCAGATCTCTGATCGAGATCTTGGCGCAAGAGGCCATCGGCAACCGTGTCTGATCGGCCCA 613
Db 2027 ACCAGATCTCTGATCGAGATCTTGGCGCAAGAGGCCATCGGCAACCGTGTCTGATCGGCCCA 2086
Qy 614 CCCCCTGTAACATCATCGGCGCCCAACATGTGTGACCCAGCTGGGTGACACCTGAACTTCC 673
Db 2087 CCCCCTGTAACATCATCGGCGCCCAACATGTGTGACCCAGCTGGGTGACACCTGAACTTCC 2146
Qy 674 CCATCAGCCCATCGACACCGTGTCCCTGTAAGCTGAAGCCCGGCATGAGCGGCCCAAGG 733
Db 2147 CCATCAGCCCATCGACACCGTGTCCCTGTAAGCTGAAGCCCGGCATGAGCGGCCCAAGG 2206
Qy 734 TGAAGCAGTGTGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCCAGGAGA 793
Db 2207 TGAAGCAGTGTGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCCAGGAGA 2266
Qy 794 TGGAGAGGAGGGCAAGATCACCAAGATCGGCGCCCGGAGAACCCCTACAACAACCCCGTGT 853
Db 2267 TGGAGAGGAGGGCAAGATCACCAAGATCGGCGCCCGGAGAACCCCTACAACAACCCCGTGT 2326
Qy 854 TCGCCATCAAGAGAGAGGAGCAGACCAAGTGTGGCGCAAGCTGTGGAGCTTTCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAGAGAGGAGCAGACCAAGTGTGGCGCAAGCTGTGGAGCTTTCGCGAGCTGA 2386
Qy 914 ACAAGCGCACCCAGGACTTCTGGAGAGTGCAGCTGGGCAATCCCCACCCCGCGCCCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGAGAGTGCAGCTGGGCAATCCCCACCCCGCGCCCTGA 2446
Qy 974 AGAAGAGAGAGAGCGTGACCGTGTGTGACGTGGGCGACGCTTCTTTCAGCGTGTGCCCTGG 1033
Db 2447 AGAAGAGAGAGAGCGTGACCGTGTGTGACGTGGGCGACGCTTCTTTCAGCGTGTGCCCTGG 2506
Qy 1034 ACAGAGACTTTCGCAAGTACACCGCTTTCACATCCCGCAGGATCAACAACGAGACCCCGG 1093
Db 2507 ACAGAGACTTTCGCAAGTACACCGCTTTCACATCCCGCAGGATCAACAACGAGACCCCGG 2566
Qy 1094 GCATCCGCTACCAAGTCAAGTGTCTGCCAGAGGCTGGAAGGGCAGCCCGCAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAAGTCAAGTGTCTGCCAGAGGCTGGAAGGGCAGCCCGCAGCATCTTCC 2626
Qy 1154 AGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCCCGCAACCCCGAGATCGTGATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCCCGCAACCCCGAGATCGTGATCT 2686
Qy 1214 ACAGGCGCCCTGTGTAGTGGGCGACACCTTGGAGATCGGCGCAGCAGCGCGCAGATCG 1273
Db 2687 ACAGGCGCCCTGTGTAGTGGGCGACACCTTGGAGATCGGCGCAGCAGCGCGCAGATCG 2746
Qy 1274 AGGAGCTGCGCAAGCAGCTGCTGGCTGGGGCTTTCACACCCCGCAGCAAGAGCACCAGA 1333
Db 2747 AGGAGCTGCGCAAGCAGCTGCTGGCTGGGGCTTTCACACCCCGCAGCAAGAGCACCAGA 2806
Qy 1334 AGGAGCCCTCTTCTCTGTGGATGGGCTTACGAGCTGCAACCCCGCAGCAAGTGGACCTGCGAGC 1393
Db 2807 AGGAGCCCTCTTCTCTGTGGATGGGCTTTCACACCCCGCAGCAAGTGGACCTGCGAGC 2860
Qy 1394 CCATCGAGCTGCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAGCTGTGTGGGCA 1453
Db 2861 CCATCGAGCTGCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAGCTGTGTGGGCA 2920
Qy 1454 AGCTGAAGTGGCGCAGCAGATCTTACCCCGGCGCATAGGTGGCGGCTGTGCAAGCTGC 1513
Db 2921 AGCTGAAGTGGCGCAGCAGATCTTACCCCGGCGCATAGGTGGCGGCTGTGCAAGCTGC 2980
```



Qy	794	TG3AGAGGAGGGCAAGATCA	CCAAAGATCGCCCCCGAGAA	CCCCCTTACAACACCCCGGT	853
Db	2267	TG3AGAGGAGGGCAAGATCA	CCAAAGATCGCCCCCGAGAA	CCCCCTTACAACACCCCGGT	2326
Qy	854	TCGCCATCAAGAAGAGCAG	CACCAAGTCGGCCCAAGCT	TGGTGGACTTCCGCGAGCTGA	913
Db	2327	TCGCCATCAAGAAGAGCAG	CACCAAGTCGGCCCAAGCT	TGGTGGACTTCCGCGAGCTGA	2386
Qy	914	ACAAGCGCACCCAGGACTT	CTTGGGAGGTGCAAGCT	CCCCCACCACCCCGCGCGCTGA	973
Db	2387	ACAAGCGCACCCAGGACTT	CTTGGGAGGTGCAAGCT	CCCCCACCACCCCGCGCGCTGA	2446
Qy	974	AGAAGAAGAGAGCGTGAC	CGTGCTGAGAGTGGGCGAG	CGCTTACTTCAGCGTGGCCCTGG	1033
Db	2447	AGAAGAAGAGAGCGTGAC	CGTGCTGAGAGTGGGCGAG	CGCTTACTTCAGCGTGGCCCTGG	2506
Qy	1034	ACGAGGACTTCCGCAAGT	PACACCGCTTACCATCCC	CAGCATCAACAACGAGACCCCG	1093
Db	2507	ACGAGGACTTCCGCAAGT	PACACCGCTTACCATCCC	CAGCATCAACAACGAGACCCCG	2566
Qy	1094	GCATCCGCTACCAAGTACA	AAAGTGTCGCCCAAGGGCT	TGAAAGGCGAGCCCGAGCATCTTCC	1153
Db	2567	GCATCCGCTACCAAGTACA	AAAGTGTCGCCCAAGGGCT	TGAAAGGCGAGCCCGAGCATCTTCC	2626
Qy	1154	AGACGAGATGACCAAGATC	CTTGAGAGCCCTTCCGGCC	CGCGAACCCTCGAGATCGTGATCT	1213
Db	2627	AGACGAGATGACCAAGATC	CTTGAGAGCCCTTCCGGCC	CGCGAACCCTCGAGATCGTGATCT	2686
Qy	1214	ACCAGGCCCCCTCTGAGT	GGGCGAGCAGCTTGGAGAT	CGGCCAGCACCGCGCCAAAGATCG	1273
Db	2687	ACCAGGCCCCCTCTGAGT	GGGCGAGCAGCTTGGAGAT	CGGCCAGCACCGCGCCAAAGATCG	2746
Qy	1274	AGGAGCTGCGCAAGCACCT	GTGCGCTGGGCTTCCAC	CCCCCGCAAGAAGCACCAGA	1333
Db	2747	AGGAGCTGCGCAAGCACCT	GTGCGCTGGGCTTCCAC	CCCCCGCAAGAAGCACCAGA	2806
Qy	1334	AGGAGCCCCCTTCTGTG	ATGGGCTACGAGCTGCA	CCCCCGCAAGTGCAGCTGCAGC	1393
Db	2807	AGGAGCCCCCTTCTGTG	ATGGGCTACGAGCTGCA	CCCCCGCAAGTGCAGCTGCAGC	2860
Qy	1394	CCATCGAGCTGCCCCGAGA	AGAGAGCTGGACCGTGA	ACGACATCCAGAAAGCTGGTGGGCA	1453
Db	2861	CCATCGAGCTGCCCCGAGA	AGAGAGCTGGACCGTGA	ACGACATCCAGAAAGCTGGTGGGCA	2920
Qy	1454	AGCTGAATGCGGCAGCGA	GTATCACCCCGCGCATCA	AGGTGCGCCAGCTGTGCAAGTGC	1513
Db	2921	AGCTGAATGCGGCAGCGA	GTATCACCCCGCGCATCA	AGGTGCGCCAGCTGTGCAAGTGC	2980
Qy	1514	TGCGGGGGCCAGGGCCCT	GTACGACATCTGTGCCCT	TGACCGGAGGCGCGAGCTGCAGC	1573
Db	2981	TGCGGGGGCCAGGGCCCT	GTACGACATCTGTGCCCT	TGACCGGAGGCGCGAGCTGCAGC	3040
Qy	1574	TGGCCGAGAACCGCGAG	ATCTCTGCGGAGCCCGT	GCACGGCGTGACTACGACCCCAAGCA	1633
Db	3041	TGGCCGAGAACCGCGAG	ATCTCTGCGGAGCCCGT	GCACGGCGTGACTACGACCCCAAGCA	3100
Qy	1634	AGGACTGTGTGGCCGAG	ATCCAGAAGCAGGGCCAC	GACCAAGTGGACCTTACAGATCTPAC	1693
Db	3101	AGGACTGTGTGGCCGAG	ATCCAGAAGCAGGGCCAC	GACCAAGTGGACCTTACAGATCTPAC	3160
Qy	1694	AGGAGCCCTTCAGAACT	GAGACCGGCAAGTAGTAC	CGCAAGATGCGCACCGCCACACCA	1753
Db	3161	AGGAGCCCTTCAGAACT	GAGACCGGCAAGTAGTAC	CGCAAGATGCGCACCGCCACACCA	3220
Qy	1754	ACGACGTGAAGCAGCTGA	CCAGGCGCGTGCAGAA	AGATCGCCATGGAGAGCATCGTGATCT	1813
Db	3221	ACGACGTGAAGCAGCTGA	CCAGGCGCGTGCAGAA	AGATCGCCATGGAGAGCATCGTGATCT	3280
Qy	1814	GGGGCAAGACCCCAAGT	TTCGCTCGCCCATTCAG	AAAGAGACCTCTGGAGACTGTGTGGA	1873
Db	3281	GGGGCAAGACCCCAAGT	TTCGCTCGCCCATTCAG	AAAGAGAGACCTCTGGAGACTGTGTGGA	3340

QY	1874	CCGACTACTGGCAGGCCACCTGGATNATCCCGAGGTGGGAGTTCTGTGAAACACCCCCCCCCCTGG	1933
Db	3341	CCGACTACTGGCAGGCCACCTGGATNATCCCGAGGTGGGAGTTCTGTGAAACACCCCCCCCCCTGG	3400
QY	1934	TGAAGCTGTGTGTACACGCTGGAGAAAGAGAGCCCATCATCGGCGCGGAGACCTTCTACGTGG	1993
Db	3401	TGAAGCTGTGTGTACACGCTGGAGAAAGAGAGCCCATCATCGGCGCGGAGACCTTCTACGTGG	3460
QY	1994	ACGGCGCGGCCAACCGCGGAGACCAAGATCGGCACGCGCGCTACGTGTACCGACCGGGGCC	2053
Db	3461	ACGGCGCGGCCAACCGCGGAGACCAAGATCGGCACGCGCGCTACGTGTACCGACCGGGGCC	3520
QY	2054	GGCAGAAAGATCGTAGGCTTACCGAGACCAACAGACCAACAGAAAGACCGAGCTGCAGGCCATCC	2113
Db	3521	GGCAGAAAGATCGTAGGCTTACCGAGACCAACAGACCAACAGAAAGACCGAGCTGCAGGCCATCC	3580
QY	2114	AGCTGGCGCCCTGCAGGACAGCGGAGCGAGGTGAACATCGTGACCGCAGCCAGCTACGCCCC	2173
Db	3581	AGCTGGCGCCCTGCAGGACAGCGGAGCGAGGTGAACATCGTGACCGCAGCCAGCTACGCCCC	3640
QY	2174	TGGGCATCATCCAGCGCCAGCCCGACAAAGAGCGAGCGAGCTGGTGAACACAGATCATCG	2233
Db	3641	TGGGCATCATCCAGCGCCAGCCCGACAAAGAGCGAGCGAGCTGGTGAACACAGATCATCG	3700
QY	2234	AGCAGCTGATCAAGAAGAGAGTGTTACTGTAGCTGGGTGCCGCCACCAAGGGGATCG	2293
Db	3701	AGCAGCTGATCAAGAAGAGAGTGTTACTGTAGCTGGGTGCCGCCACCAAGGGGATCG	3760
QY	2294	CGGCAACAGCAGCAGATCGACAAGCTGGTGAGCAAGGSCATCCGCAAGGTGCTGTTCTCTGG	2353
Db	3761	CGGCAACAGCAGCAGATCGACAAGCTGGTGAGCAAGGSCATCCGCAAGGTGCTGTTCTCTGG	3820
QY	2354	ACGGCATCGATGGCGGCATCGTGATCTACAGTACATGACGACCTGTACGTGGGAGCG	2413
Db	3821	ACGGCATCGATGGCGGCATCGTGATCTACAGTACATGACGACCTGTACGTGGGAGCG	3880
QY	2414	CGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	2457
Db	3881	CGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT	3924
RESULT 13			
US-10-190-435-58			
; Sequence 58, Application US/10190435			
; Publication No. US20030143248A1			
; GENERAL INFORMATION:			
; APPLICANT: ZUR MEGEDE, Jan			
; APPLICANT: BARNETT, Susan W.			
; APPLICANT: LIAN, Ying			
; APPLICANT: ENGELBRECHT, Susan			
; APPLICANT: VAN RENSBERG, Estrelita J.			
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C			
; FILE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF			
; FILE REFERENCE: P18133.003 / 2302-18133			
; CURRENT APPLICATION NUMBER: US/10/190,435			
; CURRENT FILING DATE: 2002-12-30			
; NUMBER OF SEQ ID NOS: 319			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 58			
; LENGTH: 5184			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefgagCpolIna			
US-10-190-435-58			

Query Match	98.0%	Score 2414;	DB 15;	Length 5184;
Best Local Similarity	99.5%;	Pred. No. 0;		
Matches 2433;	Conservative 0;	Mismatches 5;	Indels 6;	Gaps 1;
QY	14	TGGCCGAGGCCATGAGCCAGGCCACACAGCGCCCAACATCCTGTATGCAGCGCAGCAACTTCA	73	
DB	2741	TCGCCGAGGCCATGAGCCAGGCCACACAGCGCCCAACATCCTGTATGCAGCGCAGCAACTTCA	2800	

QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCCACATCGCCCGCA 133  
 Db |||||  
 2801 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCCACATCGCCCGCA 2860  
 QY 134 ACTGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGCGGCAAGGAGGCCACAGATGA 193  
 Db |||||  
 2861 ACTGCGCGCCCCCGCAAGAGGGGTGCTGGAAGTGCGGCAAGGAGGCCACAGATGA 2920  
 QY 194 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCGCGAGGACCTGGGCTTCCCGCAGGGCA 253  
 Db |||||  
 2921 AGGACTGCAACCGAGCGCGAGGCCAACTTCTTCGCGAGGACCTGGGCTTCCCGCAGGGCA 2980  
 QY 254 AGGCCCGCGAGTTCGCCAGCGAGCAGAAACCGGCCCAACAGCCCAACAGCGCGAGCTGC 313  
 Db |||||  
 2981 AGGCCCGCGAGTTCGCCAGCGAGCAGAAACCGGCCCAACAGCCCAACAGCGCGAGCTGC 3040  
 QY 314 AGGTGCGCGCGCAAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3100  
 Db |||||  
 374 TCCCCCAGATCACCTGTGGCAGCGCCCTTGGTGAGCATCAAGTGGGGGGCCAGATCA 433  
 Db |||||  
 3101 TCCCCCAGATCACCTGTGGCAGCGCCCTTGGTGAGCATCAAGTGGGGGGCGCGAGATCA 3160  
 QY 434 AGGAGGCCCTGTGGACACCGGCGCGCAGACACACCGTGTGGAGGAGATGAGCTTCCCG 493  
 Db |||||  
 3161 AGGAGGCCCTGTGGCACAACCGGCGCGCAGACACACCGTGTGGAGGAGATGAGCTTCCCG 3220  
 QY 494 GCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTAGC 553  
 Db |||||  
 3221 GCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGAGTAGC 3280  
 QY 554 ACCAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 613  
 Db |||||  
 3281 ACCAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 3340  
 QY 614 CCCCCTGGAACATCATCTGCGCGCAACATCTGACCCAGCTGGGCTGCACCTTGAATCTCC 673  
 Db |||||  
 3341 CCCCCTGGAACATCATCTGCGCGCGCAACATCTGACCCAGCTGGGCTGCACCTTGAATCTCC 3400  
 QY 674 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCAATGGAACCGCCCAAGG 733  
 Db |||||  
 3401 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCAATGGAACCGCCCAAGG 3460  
 QY 734 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCATCTGCGAGGAGA 793  
 Db |||||  
 3461 TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGGCATCTGCGAGGAGA 3520  
 QY 794 TGGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAAACCCCTTACAACACCCCGGTGT 853  
 Db |||||  
 3521 TGGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGGAGAAACCCCTTACAACACCCCGGTGT 3580  
 QY 854 TCGCCATCAAGAAGAGGAGCAGCAACCAAGTGGCGCAAGCTGTGTGACTTTCGCGAGCTGA 913  
 Db |||||  
 3581 TCGCCATCAAGAAGAGGAGCAGCAACCAAGTGGCGCAAGCTGTGTGACTTTCGCGAGCTGA 3640  
 QY 914 ACNAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGGCATCCCCACCGCGCGGCTGA 973  
 Db |||||  
 3641 ACNAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGGCATCCCCACCGCGCGGCTGA 3700  
 QY 974 AGAAGAAGAGAGCGTGCCTGCACTGGGCGAGCGCTACTTTCAGCGTGGCCCCCTGG 1033  
 Db |||||  
 3701 AGAAGAAGAGAGCGTGCCTGCACTGGGCGAGCGCTACTTTCAGCGTGGCCCCCTGG 3760  
 QY 1034 ACAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACACGAGACCCCG 1093  
 Db |||||  
 3761 ACAGGACTTTCGCAAGTACACCGCTTTCACCATCCCGCAGCATCAACACGAGACCCCG 3820  
 QY 1094 GCATCCGCTACCGATCAAGTACAGTGTCTGCCAGGCTGGAGGGCAGCCCGCAGCATCTCC 1153  
 Db |||||  
 3821 GCATCCGCTACCGATCAAGTACAGTGTCTGCCAGGCTGGAGGGCAGCCCGCAGCATCTCC 3880

QY 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGGCCCGCAACCCCGAGATCGTATCT 1213  
 Db |||||  
 3881 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGGCCCGCAACCCCGAGATCGTATCT 3940  
 QY 1214 ACAGGCCCTTCTGATCTGCTGGGAGCGACTCGAGATTCGGCCAGCACCGGGCCCAAGATCG 1273  
 Db |||||  
 3941 ACAGGCCCTTCTGATCTGCTGGGAGCGACTCGAGATTCGGCCAGCACCGGGCCCAAGATCG 4000  
 QY 1274 AGSAGCTGCGCAAGCACTTCTGCTGGGCTTCAACACCCCGCAACCAAGTGGAGCACCAGA 1333  
 Db |||||  
 4001 AGSAGCTGCGCAAGCACTTCTGCTGGGCTTCAACACCCCGCAACCAAGTGGAGCACCAGA 4060  
 QY 1334 AGSAGCCCCCTTCTCTGTGGATGAGCTACGAGCTGCACCCCGCAACCAAGTGGAGCACCAGA 1393  
 Db |||||  
 4061 AGSAGCCCCCTTCTCTGTGGCTTCTGCCCCAT-----CGAGCTGCACCCCGCAACCAAGTGGAGCACCAGA 4114  
 QY 1394 CCATCGAGCTGCCCGAGAGGAGCTGGAACCGTGAACGACATCCAGAGAGCTGGTGGCA 1453  
 Db |||||  
 4115 CCATCGAGCTGCCCGAGAGGAGCTGGAACCGTGAACGACATCCAGAGAGCTGGTGGCA 4174  
 QY 1454 AGCTGAATCTGGGCGCAGCGAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGTGC 1513  
 Db |||||  
 4175 AGCTGAATCTGGGCGCAGCGAGATCTACCCCGGATCAAGGTGCGCCAGCTGTGCAAGTGC 4234  
 QY 1514 TCGCGCGCGCCAAAGGCCCTGACCGACATCTGTCCTGACCGAGGAGGCGCGAGCTGGAGC 1573  
 Db |||||  
 4235 TCGCGCGCGCCAAAGGCCCTGACCGACATCTGTCCTGACCGAGGAGGCGCGAGCTGGAGC 4294  
 QY 1574 TGGCCGAGAAACCGCGAGATCTCTGCGGAGCCCGTGCACCGGCTGTACTAGACCCCGCA 1633  
 Db |||||  
 4295 TGGCCGAGAAACCGCGAGATCTCTGCGGAGCCCGTGCACCGGCTGTACTAGACCCCGCA 4354  
 QY 1634 AGSAGCTGCTGGCGCGAGATCCAGAGAGCGGCGCACGACAGTGGAGCTTACAGATCTACC 1693  
 Db |||||  
 4355 AGSAGCTGCTGGCGCGAGATCCAGAGAGCGGCGCACGACAGTGGAGCTTACAGATCTACC 4414  
 QY 1694 AGSAGCTTCAAGAAACCTGAAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCA 1753  
 Db |||||  
 4415 AGSAGCTTCAAGAAACCTGAAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCA 4474  
 QY 1754 AGACGTGAAGCAGCTGACCGAGGCGCTGCAGAGATTCGCCATGGAAGAGCATCGTATCT 1813  
 Db |||||  
 4475 AGACGTGAAGCAGCTGACCGAGGCGCTGCAGAGATTCGCCATGGAAGAGCATCGTATCT 4534  
 QY 1814 GGGCAAGACCCCAAGTTCCTGCTCCCATCCAGAAAGAGACCTTGGAGACCTGTGGA 1873  
 Db |||||  
 4535 GGGCAAGACCCCAAGTTCCTGCTCCCATCCAGAAAGAGACCTTGGAGACCTGTGGA 4594  
 QY 1874 CCGACTACTGGCAGGCGCACCTCGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGG 1933  
 Db |||||  
 4595 CCGACTACTGGCAGGCGCACCTCGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGG 4654  
 QY 1934 TGAAGCTGTGATACCGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTAGCTGG 1993  
 Db |||||  
 4655 TGAAGCTGTGATACCGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTAGCTGG 4714  
 QY 1994 ACGGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGAACCGAGCCGGGCC 2053  
 Db |||||  
 4715 ACGGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGAACCGAGCCGGGCC 4774  
 QY 2054 GGCAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGAGCCGAGCTGCGAGGCAATCC 2113  
 Db |||||  
 4775 GGCAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGAGCCGAGCTGCGAGGCAATCC 4834  
 QY 2114 AGCTGGCCCTTGAAGAGCAGCGGAGCGAGTGAACATCTGTGACCGAGCAGCTAGCGCC 2173  
 Db |||||  
 4835 AGCTGGCCCTTGAAGAGCAGCGGAGCGAGTGAACATCTGTGACCGAGCAGCTAGCGCC 4894  
 QY 2174 TGGGCAATCATCAGAGCCCGAGCCGCAAGAGCGAGCGAGCTGGTGAACAGATCATCG 2233  
 Db |||||  
 4895 TGGGCAATCATCAGAGCCCGAGCCCGCAAGAGCGAGCGAGCTGGTGAACAGATCATCG 4954  
 QY 2234 AGCAGCTGATCAAGAGAGGAGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCG 2293





4235	Db		TGCGCGCGCCAAAGCCCTGACGACATCTGTGCCCCCTGACCGAGGAGCGCGAGCTGTGAGC	4239
1574	Qy		TGGCCGAGAAACCGCGAGATCTTGCGCAGAGCCCGTGACGCGCGTGTACTACGACCCCCAGCA	1633
4295	Db		TGGCCGAGAACCGCGAGATCTTGCGCAGAGCCCGTGACGCGGTGTACTACGACCCCGACGA	4354
1634	Qy		AGGACTTGTTGGCCGAGATCTAGAAAGCAGAGGCGACGACGAGTGGACCTTACAGATCTTACC	1693
4355	Db		AGGACTTGTTGGCCGAGATCTAGAAAGCAGAGGCGCCACGACGAGTGGACCTTACAGATCTTACC	4414
1694	Qy		AGGAGCCCTTTCAGAACTCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCA	1753
4415	Db		AGGAGCCCTTTCAGAACTCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCA	4474
1754	Qy		ACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGGAGAGCATCGTGTATCT	1813
4475	Db		ACGACGTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGGAGAGCATCGTGTATCT	4534
1814	Qy		GGGGCAAGACCCCAAGTTCCGCTGCCATTCAGAAAGGAGACTCTGGGAGACCTTGTGTGA	1873
4535	Db		GGGGCAAGACCCCAAGTTCCGCTGCCATTCAGAAAGGAGACTCTGGGAGACCTTGTGTGA	4594
1874	Qy		CCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCCCTCTGG	1933
4595	Db		CCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCCCTCTGG	4654
1934	Qy		TGAAGCTGTGGTACCACTGCGAGAGAGGCCATCATCTCGCGCCGAGACCTTCTACGTGG	1993
4655	Db		TGAAGCTGTGGTACCACTGCGAGAGAGGCCATCATCTCGCGCCGAGACCTTCTACGTGG	4714
1994	Qy		ACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGGGGCC	2053
4715	Db		ACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGGGGCC	4774
2054	Qy		GGCAGAGATCTGTAGCCTGACCGAGACACCAACACAGAGAGCCGAGCTCGAGGCCATCC	2113
4775	Db		GGCAGAGATCTGTAGCCTGACCGAGACACCAACACAGAGAGCCGAGCTCGAGGCCATCC	4834
2114	Qy		AGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCC	2173
4835	Db		AGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCC	4894
2174	Qy		TGGGCATCATCCAGGCCAGCCCGACAGAGCGAGCGAGCTGGTGAACACAGATCATCG	2233
4895	Db		TGGGCATCATCCAGGCCAGCCCGACAGAGCGAGCGAGCTGGTGAACACAGATCATCG	4954
2234	Qy		AGCAGCTGATCAGAAGAGAGAGTGTACTCTGAGCTGGTGCCTCCGCCACAGAGGAGTCG	2293
4955	Db		AGCAGCTGATCAGAAGAGAGAGTGTACTCTGAGCTGGTGCCTCCGCCACAGAGGAGTCG	5014
2294	Qy		CGGSCAACGAGCAGATCGCAAGCTGTGTAGCAAGGGCATCCGCAAGTGTCTGTCTCTGG	2353
5015	Db		CGGSCAACGAGCAGATCGCAAGCTGTGTAGCAAGGGCATCCGCAAGTGTCTGTCTCTGG	5074
2354	Qy		ACGGCATCGATGGCGGCATCGTGATCTTACAGTACATGGACGACCTGTAGCTGGGACGC	2413
5075	Db		ACGGCATCGATGGCGGCATCGTGATCTTACAGTACATGGACGACCTGTAGCTGGGACGC	5134
2414	Qy		CGCGCCCTTAGATCGATTAAAGTTCCCGGGCTTAGCACCGGT	2457
5135	Db		CGCGCCCTTAGATCGATTAAAGTTCCCGGGCTTAGCACCGGT	5178

RESULT 15

US-10-190-435-13

00 10 130 133 135  
; Sequence 13, Application US/10190435

; Publication No. US20030143248A1

GENERAL INFORMATION:

APPLICANT: ZUR MEGEDE, Jan

APPLICANT: BARNETT, Susan W.

; APPLICANT: LIAN, Ying

; APPLICANT: ENGELBRECHT, Susan

APPLICANT: VAN RENSBURG, Estrelita J.

```

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
;
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
;
; FILE REFERENCE: PP18133.003 / 2302-18133
;
; CURRENT APPLICATION NUMBER: US/10/190,435
;
; CURRENT FILING DATE: 2002-12-30
;
; NUMBER OF SEQ ID NOS: 319
;
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ. ID. NO.13
;
; LENGTH: 3531
;
; TYPE: DNA
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut_C
US-10-190-435-13

```

Query Match	96.8%	Score 2383.6	DB 15	Length 3531
Best Local Similarity	98.8%	Pred. No. 0		
Matches 2414	Conservative 0	Mismatches 24	Indels 6	Gaps 1
Qy	14	TGGCCGAGGCATGAGCCAGGCCACACAGAGCCCAACATCCTGATGCGAGCGAGCAACTTCA	73	
Db	1088	TGGCCGAGGCATGAGCCAGGCCCAACACAGCGTGATGATGCAAGAGCACTTTAAAA	1147	
Qy	74	AGGCGCCCAAGCGCATCATCAAGTGTCTTAACTGCGGCAAGGAGGGCCACATCGCCGCA	133	
Db	1148	AGGCGCCCAAGCGCATCATCAAGTGTCTTAACTGCGGCAAGGAGGGCCACATCGCCGCA	1207	
Qy	134	ACTGCGCGCCCCCGGAGAGAGGGTGTCTGGAAGTTCGGCAAGGAGGGCCACCATGATGA	193	
Db	1208	ACTGCGCGCCCCCGGAGAGAGGGTGTCTGGAAGTTCGGCAAGGAGGGCCACCATGATGA	126	
Qy	194	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCTTCTCCCCCAGGGCA	253	
Db	1268	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTGGCTTCTCCCCCAGGGCA	1327	
Qy	254	AGGCGCGGAGTTCGCCAGGAGCAGAAACCGGCCCAACAGCCCCAACAGCGCGAGCTGC	313	
Db	1328	AGGCGCGGAGTTCGCCAGGAGCAGAAACCGGCCCAACAGCCCCAACAGCGCGAGCTGC	1387	
Qy	314	AGGTGCGCGGAGCAACCCCGCAGCGAGCGCCGGCGCGAGCGCGGAGGACACCTGAACT	373	
Db	1388	AGGTGCGCGGAGCAACCCCGCAGCGAGCGCCGGCGCGAGCGCGGAGGACACCTGAACT	1447	
Qy	374	TCCCCCAGATCAACCTGTGTGACGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGCAGATCA	433	
Db	1448	TCCCCCAGATCAACCTGTGTGACGCGCCCTGTGTGAGCATCAAGGTGGGCGGCGCAGATCA	1507	
Qy	434	AGGAGGCCCTGTGTGACACCGGCGCCGACACACCGTGTGTGGAGGAGATGAGCCTGC	493	
Db	1508	AGGAGGCCCTGTGTGACACCGGCGCCGACACACCGTGTGTGGAGGAGATGAGCCTGC	1567	
Qy	494	GCAAGTGGAGGCCAAGATGATCGGGGCGATCGCGGCTTCATCAAGGTGGCGCAGTACG	553	
Db	1568	GCAAGTGGAGGCCAAGATGATCGGGGCGATCGCGGCTTCATCAAGGTGGCGCAGTACG	1627	
Qy	554	ACCAGATCTGTGATGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA	613	
Db	1628	ACCAGATCTGTGATGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCA	1687	
Qy	614	CCCCCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCCCTGAACTTC	673	
Db	1688	CCCCCGTGAACATCATCGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCCCTGAACTTC	1747	
Qy	674	CCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCGATGGAAGGCCCCAAGG	733	
Db	1748	CCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCGATGGAAGGCCCCAAGG	1807	
Qy	734	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA	793	
Db	1808	TGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA	1867	
Qy	794	TGGAGAGGAGGGCAAGATCACCAAGATCGGCCCGGAGAAACCCCTTACAAACACCCCGTGT	853	



Db 1868 TGGAGAGGAGGGCAAGATCACCAAGATCGGCGCCGAGAACCCCTACACACCCCGGTGT 1927  
Qy 854 TCGCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 913  
Db 1928 TCGCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 1987  
Qy 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGCTGGGCAATCCCCACCCCGCGGCTGA 973  
Db 1988 ACAAGCGCACCCAGGACTTCTGGGAGGTGCAAGCTGGGCAATCCCCACCCCGCGGCTGA 2047  
Qy 974 AGAAGAAAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTTCTTACGCGTGGCCCTGG 1033  
Db 2048 AGAAGAAAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTTCTTACGCGTGGCCCTGG 2107  
Qy 1034 ACGAGGACTTCCGCAAGTACACCCGCTTCAACCATCCCGAGCATCAACAAGAGACCCCG 1093  
Db 2108 ACGAGGACTTCCGCAAGTACACCCGCTTCAACCATCCCGAGCATCAACAAGAGACCCCG 2167  
Qy 1094 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGCTGGAGGGCAGCCCCAGCATCTCC 1153  
Db 2168 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGCTGGAGGGCAGCCCCAGCATCTCC 2227  
Qy 1154 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCT 1213  
Db 2228 AGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCT 2287  
Qy 1214 ACCAGGCCCTTGTACGTGGGCGAGCAGCTGGAGATCGGCGCAGCACCGCGCCCAAGATCG 1273  
Db 2288 ACCAGGCCCTTGTACGTGGGCGAGCAGCTGGAGATCGGCGCAGCACCGCGCCCAAGATCG 2347  
Qy 1274 AGGAGCTGGCGAGCAGCTGCTGGGCTGGGCTTCAACACCCCGCAGCAAGAGACACAGA 1333  
Db 2348 AGGAGCTGGCGAGCAGCTGCTGGGCTGGGCTTCAACACCCCGCAGCAAGAGACACAGA 2407  
Qy 1334 AGGAGCCCTTCTGTGTGATGGCTACGAGCTGCAACCCGCAAGTGGACCGTGGAGC 1393  
Db 2408 AGGAGCCCTTCTGTGTGATGGCTACGAGCTGCAACCCGCAAGTGGACCGTGGAGC 2461  
Qy 1394 CCATCGAGCTGCCGAGAGGAGAGCTGGACCGGTGAACGACATCCAGAAAGCTGGTGGCA 1453  
Db 2462 CCATCGAGCTGCCGAGAGGAGAGCTGGACCGGTGAACGACATCCAGAAAGCTGGTGGCA 2521  
Qy 1454 AGCTGAATGGGCGAGCAGATCTACCCCGGCAATCAAGGTGGCGCAGCTGTGCAAGCTGC 1513  
Db 2522 AGCTGAATGGGCGAGCAGATCTACCCCGGCAATCAAGGTGGCGCAGCTGTGCAAGCTGC 2581  
Qy 1514 TGGCGGCGCCAGGCGCTGACCGACATCGTGCCCTGACCGAGGAGCGCGAGCTGGAGC 1573  
Db 2582 TGGCGGCGCCAGGCGCTGACCGACATCGTGCCCTGACCGAGGAGCGCGAGCTGGAGC 2641  
Qy 1574 TGGCGGAGAACCGGAGATCTGCGCGAGCCGCTGCAAGCTGTACTACGACCCGAGCA 1633  
Db 2642 TGGCGGAGAACCGGAGATCTGCGCGAGCCGCTGCAAGCTGTACTACGACCCGAGCA 2701  
Qy 1634 AGGACCTGGTGGCGAGATCCAGAGCAGAGGCGCCACGACAGTGGACCTTACAGATCTACC 1693  
Db 2702 AGGACCTGGTGGCGAGATCCAGAGCAGAGGCGCCACGACAGTGGACCTTACAGATCTACC 2761  
Qy 1694 AGGAGCCCTTCAAGAACCTGAGACCGGCAAGTACGCAAGATCGCGACCCGACCA 1753  
Db 2762 AGGAGCCCTTCAAGAACCTGAGAACCGGCAAGTACGCAAGATCGCGACCCGACCA 2821  
Qy 1754 ACGACGTGAAGAGCAGCTGACCGAGGCGCTGCAAGATCGGCATGGAGAGCATCGTGATCT 1813  
Db 2822 ACGACGTGAAGAGCAGCTGACCGAGGCGCTGCAAGATCGGCATGGAGAGCATCGTGATCT 2881  
Qy 1814 GGGGCAAGACCCCAAGTTCCGCTGCGCTTCCATCCAGAAAGGAGACCTTGGGAGACCTTGGTGA 1873  
Db 2882 GGGGCAAGACCCCAAGTTCCGCTGCGCTTCCATCCAGAAAGGAGACCTTGGGAGACCTTGGTGA 2941  
Qy 1874 CCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG 1933  
Db 2942 CCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG 3001

Qy 1934 TGAAGCTGTGGTACCAAGCTGGAGAGGAGGCCCATCATCGGCGCGAGACCTTCTTACGTGG 1993  
Db 3002 TGAAGCTGTGGTACCAAGCTGGAGAGGAGGCCCATCATCGGCGCGAGACCTTCTTACGTGG 3061  
Qy 1994 ACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGGGCC 2053  
Db 3062 ACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGGGCC 3121  
Qy 2054 GGCAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAACCGAGCTGCAGGGCATCC 2113  
Db 3122 GGCAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAACCGAGCTGCAGGGCATCC 3181  
Qy 2114 AGCTGGCCCTGACGACAGCGGCGAGCGAGGTGAACATGTGAACGACAGCCAGTACGCC 2173  
Db 3182 AGCTGGCCCTGACGACAGCGGCGAGCGAGGTGAACATGTGAACGACAGCCAGTACGCC 3241  
Qy 2174 TGGGCATCATCCAGGCCCGCCGACAGAGCGAGGAGCTGGTGAACAGATCATCG 2233  
Db 3242 TGGGCATCATCCAGGCCCGCCGACAGAGCGAGGAGCTGGTGAACAGATCATCG 3301  
Qy 2234 AGCAGCTGATCAAGAAAGAGAGGTGTACTGAGCTGGGTGCCCGCCACCAAGGGCATCG 2293  
Db 3302 AGCAGCTGATCAAGAAAGAGAGGTGTACTGAGCTGGGTGCCCGCCACCAAGGGCATCG 3361  
Qy 2294 GCGCAACGAGCAGATCGACAAAGCTGGTGAACAAAGGGCATCCGCAAGGTGCTGTCTTGG 2353  
Db 3362 GCGCAACGAGCAGATCGACAAAGCTGGTGAACAAAGGGCATCCGCAAGGTGCTGTCTTGG 3421  
Qy 2354 ACGGCATCGATGGCGGCGATCGTGATCTACCGATACATGGAGACCTGTACGTGGGCGAG 2413  
Db 3422 ACGGCATCGATGGCGGCGATCGTGATCTACCGATACATGGAGACCTGTACGTGGGCGAG 3481  
Qy 2414 GCGGCCCTAGGATCGATTAAGAGCTTCCCGGGGCTAGCACCGGT 2457  
Db 3482 GCGGCCCTAGGATCGATTAAGAGCTTCCCGGGGCTAGCACCGGT 3525

Search completed: June 2, 2005, 04:23:40  
Job time : 1419.33 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 03:50:55 ; Search time 10513 Seconds  
(without alignments)  
11324.506 Million cell updates/sec

Title: US-09-610-313B-32  
Perfect score: 2457  
Sequence: 1 gtcagccaccatggccga.....gggctagaccggtgaattc 2457

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_ets.\*
- 12: gb\_ey.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	100.0	2457	6	AX455916
2	2436.2	99.2	2463	6	AX455915 Sequence
3	2415.4	98.3	2469	6	AX455914 Sequence
4	2040	83.0	2300	6	BD263705 Improved
5	2040	83.0	2300	6	CQ870575 Sequence
6	2040	83.0	2300	6	AR373388 Sequence
7	2019.2	82.2	2306	6	BD263704 Improved
8	2019.2	82.2	2306	6	CQ870574 Sequence
9	2019.2	82.2	2306	6	AR373387 Sequence
10	2000.8	81.4	9165	6	AX427930 Sequence
11	1998.4	81.3	2312	6	BD263706 Improved
12	1998.4	81.3	2312	6	CQ870576 Sequence
13	1998.4	81.3	2312	6	AR373389 Sequence
14	1992.4	81.1	9788	6	AX427936 Sequence
15	1978.4	80.5	9169	6	AX427931 Sequence
16	1975	80.4	9194	6	AX427926 Sequence
17	1973.4	80.3	9194	6	AX427925 Sequence
18	1973.4	80.3	12411	6	AX427927 Sequence
19	1970	80.2	9785	6	AX427938 Sequence

20	1966.6	80.0	9167	6	AX427933	AX427933 Sequence
21	1966.6	80.0	9170	6	AX427928	AX427928 Sequence
22	1966.6	80.0	9782	6	AX427935	AX427935 Sequence
23	1966.6	80.0	9783	6	AX427934	AX427934 Sequence
24	1966.6	80.0	9782	6	AX427932	AX427932 Sequence
25	1965	80.0	9189	6	AX427921	AX427921 Sequence
26	1962	79.9	4353	12	AF287353	AF287353 Synthetic
27	1945.4	79.2	3015	6	AX455946	AX455946 Sequence
28	1928.8	78.5	3009	6	AX455987	AX455987 Sequence
29	1915.8	78.0	4319	6	BD263639	BD263639 Improved
30	1915.8	78.0	4319	6	CQ870498	CQ870498 Sequence
31	1915.8	78.0	4319	6	AR373322	AR373322 Sequence
32	1905.2	77.5	4352	12	AF287352	AF287352 Synthetic
33	1878.6	76.5	1965	6	AX455952	AX455952 Sequence
34	1872.8	76.2	2299	6	BD263703	BD263703 Improved
35	1872.8	76.2	2299	6	CQ870573	CQ870573 Sequence
36	1872.8	76.2	2299	6	AR373386	AR373386 Sequence
37	1859.8	75.7	2577	6	AX457088	AX457088 Sequence
38	1852.2	75.4	4343	6	AX306428	AX306428 Sequence
39	1852	75.4	2305	6	BD263702	BD263702 Improved
40	1852	75.4	2305	6	CQ870572	CQ870572 Sequence
41	1852	75.4	2305	6	AR373385	AR373385 Sequence
42	1848.2	75.2	4341	6	AX306429	AX306429 Sequence
43	1838.6	74.8	1977	6	AX455954	AX455954 Sequence
44	1825.6	74.3	7897	6	AX427923	AX427923 Sequence
45	1772	72.1	3024	12	AF287355	AF287355 Synthetic

ALIGNMENTS

RESULT 1  
AX455916  
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 32 from Patent WO0204493.  
ACCESSION AX455916  
VERSION AX455916.1 GI:21714901  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 32 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
Location/Qualifiers  
source  
1..2457  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YMW"

ORIGIN

Query Match 100.0%; Score 2457; DB 6; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 1.5e-247;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTCAGCCCAACCATGGCCGAGGCATGAGCGCCACCGCCCAATCCTCTGATGAG 60  
Db 1 GTCAGCCCAACCATGGCCGAGGCATGAGCGCCACCGCCCAATCCTCTGATGAG 60  
Qy 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTCTCACTCGGCAAGGCGC 120  
Db 61 CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTCTCACTCGGCAAGGCGC 120  
Qy 121 CACATCGCCGCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCAAGGAG 180  
Db 121 CACATCGCCGCAACTGCGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCAAGGAG 180  
Qy 181 GGCCACCAAGTAGAGGACTGCACCGAGCGCCCAACTTCTTCGCGAGGAGCTCGCC 240  
Db 181 GGCCACCAAGTAGAGGACTGCACCGAGCGCCCAACTTCTTCGCGAGGAGCTCGCC 240

Db 181 GGCCACCAAGATGAAGACTGCAACGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCC 240  
Qy 241 TTCCCCCGAGGCAAGCGCGAGTTCCCGAGGAGCAAGACCGCGCAAGAGCCCAACC 300  
Db 241 TTCCCCCGAGGCAAGCGCGAGTTCCCGAGGAGCAAGACCGCGCAAGAGCCCAACC 300  
Qy 301 AGCCCGAGGCTCAGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGAGCGCCAG 360  
Db 301 AGCCCGAGGCTCAGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGAGCGCCAG 360  
Qy 361 GGCAACCTGAACTTCCCGAGATCAACCTGTGGCAGCGCCCTTGGTGAAGATCAAGGTG 420  
Db 361 GGCAACCTGAACTTCCCGAGATCAACCTGTGGCAGCGCCCTTGGTGAAGATCAAGGTG 420  
Qy 421 GGCGGCCAGATCAAGAGGCCCTGTGGACACCGCGCGCGAGCAACCGTGTGGAGGAG 480  
Db 421 GGCGGCCAGATCAAGAGGCCCTGTGGACACCGCGCGCGAGCAACCGTGTGGAGGAG 480  
Qy 481 ATGAGCTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGGCTTCAATCAAG 540  
Db 481 ATGAGCTGCGCGCAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGGCTTCAATCAAG 540  
Qy 541 GTGCGCCAGTACGACAGATCTGATCGAGATCTCGGCAAGAGGCCATCGGCAACCGTG 600  
Db 541 GTGCGCCAGTACGACAGATCTGATCGAGATCTCGGCAAGAGGCCATCGGCAACCGTG 600  
Qy 601 CTGATCGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Qy 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
Qy 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCC 780  
Db 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCC 780  
Qy 781 ATCTGGAGGAGATGGAGAGGCGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
Db 781 ATCTGGAGGAGATGGAGAGGCGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
Qy 841 AACACCCCGCTGTGCCATCAAGAAAGAGCAGCACCAAGTGGCGAAAGCTGGTGGAC 900  
Db 841 AACACCCCGCTGTGCCATCAAGAAAGAGCAGCACCAAGTGGCGAAAGCTGGTGGAC 900  
Qy 901 TTCCCGAGCTGAAACAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCATCCCCAC 960  
Db 901 TTCCCGAGCTGAAACAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCATCCCCAC 960  
Qy 961 CCGCGCGGCTGAAGAGAGAGCGTGACCGTGTGGACGTGGGCGACGCTACTTTC 1020  
Db 961 CCGCGCGGCTGAAGAGAGAGCGTGACCGTGTGGACGTGGGCGACGCTACTTTC 1020  
Qy 1021 AGCTGCCCCCTGAGAGGACTTCGCAAGTACACCGCTTCCACATCCCCAGCATCAAC 1080  
Db 1021 AGCTGCCCCCTGAGAGGACTTCGCAAGTACACCGCTTCCACATCCCCAGCATCAAC 1080  
Qy 1081 AACAGAGACCCCGGCATCCGCTACAGTACAACGTGTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
Db 1081 AACAGAGACCCCGGCATCCGCTACAGTACAACGTGTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
Qy 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTTGAGCCCTTTCGGCGCCCGCAACCCC 1200  
Db 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTTGAGCCCTTTCGGCGCCCGCAACCCC 1200  
Qy 1201 GAGATCGTATCTACAGGCCCCCTGTACGTGGGCGAGCGACTTGAGATCGGCGCAGCAGC 1260  
Db 1201 GAGATCGTATCTACAGGCCCCCTGTACGTGGGCGAGCGACTTGAGATCGGCGCAGCAGC 1260  
Qy 1261 CGCGCCAAAGATCGAGAGCTGCGCAAGCACTGCTGCGTGGGGCTTCAACACCCCGGAC 1320  
Db 1261 CGCGCCAAAGATCGAGAGCTGCGCAAGCACTGCTGCGTGGGGCTTCAACACCCCGGAC 1320

Qy 1321 AAGAAGCACCAAGAGAGCCCTTCTCTGCCATCGAGCTGCACCCCGACAAGTGGACC 1380  
Db 1321 AAGAAGCACCAAGAGAGCCCTTCTCTGCCATCGAGCTGCACCCCGACAAGTGGACC 1380  
Qy 1381 GTGCAAGCCATCGAGCTGCGCGAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTG 1440  
Db 1381 GTGCAAGCCATCGAGCTGCGCGAGAGAGAGTGGACCGTGAACGACATCCAGAAGCTG 1440  
Qy 1441 GTGGCAAGCTGAATGGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGC 1500  
Db 1441 GTGGCAAGCTGAATGGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGC 1500  
Qy 1501 AAGTGTCTGCGCGCGCAAGCCCTGACCGACATCTGTGCCCTCACCGAGAGGCCGAG 1560  
Db 1501 AAGTGTCTGCGCGCGCAAGCCCTGACCGACATCTGTGCCCTCACCGAGAGGCCGAG 1560  
Qy 1561 CTGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCAAGCTGTACTACGAC 1620  
Db 1561 CTGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCAAGCTGTACTACGAC 1620  
Qy 1621 CCCAGAAAGACCTGGTGGCCGAGATCCAGAAAGCGAGGCCACGACAGTGGACCTACCG 1680  
Db 1621 CCCAGAAAGACCTGGTGGCCGAGATCCAGAAAGCGAGGCCACGACAGTGGACCTACCG 1680  
Qy 1681 ATCTACCAAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1740  
Db 1681 ATCTACCAAGGACCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1740  
Qy 1741 CACACCAAGCAGTGAAGCAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAGAGCATC 1800  
Db 1741 CACACCAAGCAGTGAAGCAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAGAGCATC 1800  
Qy 1801 GTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGACCTGGAGAGACC 1860  
Db 1801 GTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGACCTGGAGAGACC 1860  
Qy 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCTGTGAACACCCC 1920  
Db 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCTGTGAACACCCC 1920  
Qy 1921 CCCCTGTGAAGCTGTGTGTACAGTGGAAGAGGCCCATCATCGGCGCGAGACCTTTC 1980  
Db 1921 CCCCTGTGAAGCTGTGTGTACAGTGGAAGAGGCCCATCATCGGCGCGAGACCTTTC 1980  
Qy 1981 TACGTGGAAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040  
Db 1981 TACGTGGAAGCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGAC 2040  
Qy 2041 CGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2100  
Db 2041 CGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2100  
Qy 2101 GCCATCCAGCTGGCCCTGCAGGACAGCGCGAGCGAGGTGAACATCGTGACCGAGCCAG 2160  
Db 2101 GCCATCCAGCTGGCCCTGCAGGACAGCGCGAGCGAGGTGAACATCGTGACCGAGCCAG 2160  
Qy 2161 TACGCCCTGGGCATCATCCAGGCCCGAGCAAGAGCGAGCGAGCTGTGTGAACCGAG 2220  
Db 2161 TACGCCCTGGGCATCATCCAGGCCCGAGCAAGAGCGAGCGAGCTGTGTGAACCGAG 2220  
Qy 2221 ATCATCGAGCAGCTGATCAAGAGAGAGGTGTACCTGAGCTGGTGGCCCGCCACAG 2280  
Db 2221 ATCATCGAGCAGCTGATCAAGAGAGAGGTGTACCTGAGCTGGTGGCCCGCCACAG 2280  
Qy 2281 GGCATCGGCGCAACAGCAGATCGAAGCTGGTGGAGCAAGGGGCATCCGCAAGGTGTG 2340  
Db 2281 GGCATCGGCGCAACAGCAGATCGAAGCTGGTGGAGCAAGGGGCATCCGCAAGGTGTG 2340  
Qy 2341 TTCTCTGAGCGGCATCGATGGGGCATCGTATCTACAGTACATGAGACCTGTACCTG 2400  
Db 2341 TTCTCTGAGCGGCATCGATGGGGCATCGTATCTACAGTACATGAGACCTGTACCTG 2400

Qy 2401 GGCAGCGCGCCCTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
Db |||||||  
2401 GGCAGCGCGCCCTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
|||

RESULT 2  
AX455915  
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 31 from Patent WO0204493.  
ACCESSION AX455915  
VERSION AX455915.1 GI:21714900  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
other sequences; artificial sequences.

REFERENCE 1  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Renaburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
Location/Qualifiers  
1..2463  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975YM"

ORIGIN  
Query Match 99.2%; Score 2436.2; DB 6; Length 2463;  
Best Local Similarity 99.6%; Pred. No. 2.1e-245;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
1; 1;  
Qy 1 GTCAGCGCCACATGCGCGAGCGCATGAGCGAGCGCCAGCGCCAGCATCTCTGATGCG 60  
Db |||||||  
1 GTCAGCGCCACATGCGCGAGCGCATGAGCGAGCGCCAGCGCCAGCATCTCTGATGCG 60  
|||

Qy 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGGGC 120  
Db |||||||  
61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGGGC 120  
|||

Qy 121 CACATCGCGCGCAACTGCGCGGCCCCCGCGAAGAGGGCTGCTGGAGAGTGGCGCAAGGAG 180  
Db |||||||  
121 CACATCGCGCGCAACTGCGCGGCCCCCGCGAAGAGGGCTGCTGGAGAGTGGCGCAAGGAG 180  
|||

Qy 181 GGCCACACAGATGAAGAGTGCACAGCGCGCGAGCGCAACTTCTCCGCGAGGACTGGCC 240  
Db |||||||  
181 GGCCACACAGATGAAGAGTGCACAGCGCGCGAGCGCAACTTCTCCGCGAGGACTGGCC 240  
|||

Qy 241 TTCCCCCGAGGGCAAGGGCCGCGAGTTCCTCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300  
Db |||||||  
241 TTCCCCCGAGGGCAAGGGCCGCGAGTTCCTCCAGCGAGCAGAACCGCGCCCAACAGCCCCACC 300  
|||

Qy 301 AGCGCGAGCTGCGAGTGGCGGCGACAAACCCCGCGAGAGGCGCGGCGCGAGCGCGAG 360  
Db |||||||  
301 AGCGCGAGCTGCGAGTGGCGGCGACAAACCCCGCGAGAGGCGCGGCGCGAGCGCGAG 360  
|||

Qy 361 GGCAACCTGAATCTTCCCGCGAGATCAACCTGTCGCGAGCGCGCCCTGCTGAGCATCAAGGTG 420  
Db |||||||  
361 GGCAACCTGAATCTTCCCGCGAGATCAACCTGTCGCGAGCGCGCCCTGCTGAGCATCAAGGTG 420  
|||

Qy 421 GGCGGCGAGATCAAGGAGGGCCCTGCTGGACACCGCGCGCGAGCACACCGTGTGGAGGAG 480  
Db |||||||  
421 GGCGGCGAGATCAAGGAGGGCCCTGCTGGACACCGCGCGCGAGCACACCGTGTGGAGGAG 480  
|||

Qy 481 ATGAGCTGCGCGCGAGTGAAGCGCCAAAGATGATGCGCGCGCTTCAATCAAG 540  
Db |||||||  
481 ATGAGCTGCGCGCGAGTGAAGCGCCAAAGATGATGCGCGCGCTTCAATCAAG 540  
|||

Qy 541 GTGGCGGAGTACGACAGATCTGATCGAGATCTGCGGCAAGAGGGCCATCGGACCGGTG 600  
Db |||||||  
541 GTGGCGGAGTACGACAGATCTGATCGAGATCTGCGGCAAGAGGGCCATCGGACCGGTG 600  
|||

Qy 601 CTGATCGCGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGTGAACCCAGCTGGGGTGC 660  
Db |||||||  
601 CTGATCGCGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGTGAACCCAGCTGGGGTGC 660  
|||

Qy 661 ACCCTGAACCTTCCCATCAGCCCATCGAGACCGTCCCGGTGAAGCTGAAGCCCGGATG 720  
Db |||||||  
661 ACCCTGAACCTTCCCATCAGCCCATCGAGACCGTCCCGGTGAAGCTGAAGCCCGGATG 720  
|||

Qy 721 GACCGGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTTGAACGCC 780  
Db |||||||  
721 GACCGGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTTGAACGCC 780  
|||

Qy 781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGCGCCCGAGAACCCCTTAC 840  
Db |||||||  
781 ATCTGCGAGGAGATGGAAGAGGAGGCAAGATCAACCAAGATCGCGCCCGAGAACCCCTTAC 840  
|||

Qy 841 AACACCCCGCTTTCGCATCAAG 900  
Db |||||||  
841 AACACCCCGCTTTCGCATCAAG 900  
|||

Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGACAGTGGGATCCCCAC 960  
Db |||||||  
901 TTCCGCGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGACAGTGGGATCCCCAC 960  
|||

Qy 961 CCGCGCGCTGAAG 1020  
Db |||||||  
961 CCGCGCGCTGAAG 1020  
|||

Qy 1021 AGCGTGGCCCTGAG 1080  
Db |||||||  
1021 AGCGTGGCCCTGAG 1080  
|||

Qy 1081 AACGAGAGACCCCGGCTACCGCTACAGTACAACTGTGCTGCCCGAGGGCTGAAGGGGAGC 1140  
Db |||||||  
1081 AACGAGAGACCCCGGCTACCGCTACAGTACAACTGTGCTGCCCGAGGGCTGAAGGGGAGC 1140  
|||

Qy 1141 CCGAGCATCTTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCGCCCAACCC 1200  
Db |||||||  
1141 CCGAGCATCTTCCAGAGCAGATGACCAAGATCTTGGAGCCCTTCCGCGCGCCCAACCC 1200  
|||

Qy 1201 GAGATCTGATCTACAGGCGCCCTGTAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
Db |||||||  
1201 GAGATCTGATCTACAGGCGCCCTGTAGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
|||

Qy 1261 CGCGCCAAAGATCGAGGAGCTGCGAAGCAGCTGCTGCGTGGGGCTTCCACCAACCCCGAC 1320  
Db |||||||  
1261 CGCGCCAAAGATCGAGGAGCTGCGAAGCAGCTGCTGCGTGGGGCTTCCACCAACCCCGAC 1320  
|||

Qy 1321 AAGAAGCACAG 1374  
Db |||||||  
1321 AAGAAGCACAG 1374  
|||

Qy 1375 TGAGACCGTGAAGCCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434  
Db |||||||  
1375 TGAGACCGTGAAGCCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434  
|||

Qy 1435 AAGCTGTGGGCAAGCTGAAGCTGGCGCAGCAGATCTACCCCGCAGATCAAGGTCGCGCAG 1494  
Db |||||||  
1435 AAGCTGTGGGCAAGCTGAAGCTGGCGCAGCAGATCTACCCCGCAGATCAAGGTCGCGCAG 1494  
|||

Qy 1495 CTGTGAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCTGCGCCCTGACCGAGAGAG 1554  
Db |||||||  
1495 CTGTGAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCTGCGCCCTGACCGAGAGAG 1554  
|||

Qy 1555 GCGAGCTGAGCTGCGCGAGAGAACCGCGAGATCTGCGCGAGAGAGAGAGAGAGAGAGAG 1614  
Db |||||||  
1555 GCGAGCTGAGCTGCGCGAGAGAACCGCGAGATCTGCGCGAGAGAGAGAGAGAGAGAGAG 1614  
|||

Qy 1615 TAGGACCCAG 1674  
Db |||||||  
1615 TAGGACCCAG 1674  
|||

Qy 1675 TACGAGATCTACGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGATGCGC 1734  
Db |||||||  
1675 TACGAGATCTACGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGATGCGC 1734  
|||

```
|||||
1681 TACCAGATCTACCAAGAGCCCTTCAAGAACCTTGAAGACGGCAAGTACGCCAAGATGCGC 1740
QY ACCGCCCAACCAACGACGTAAGAGAGTGACCGAGGGCGTGCGAAGATCGCCATGGAG 1794
Db ACCGCCCAACCAACGACGTAAGAGAGTGACCGAGGGCGGTGCGAAGATCGCCATGGAG 1800
QY AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGG 1854
Db AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGG 1860
QY GAGACCTGGTGACCGACCTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTGAAC 1914
Db GAGACCTGGTGACCGACCTACTTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCTGTGAAC 1920
QY ACCCCCCCTGGTGAAGCTGTGTGACAGCTGTGAGAGAGAGGCCCATCATCGGGCCGAG 1974
Db ACCCCCCCTGGTGAAGCTGTGTGACAGCTGTGAGAGAGAGGCCCATCATCGGGCCGAG 1980
QY ACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTACCTG 2034
Db ACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGGCTACCTG 2040
QY ACCGACCGGGCGCGCAGAGAATCGTGAGCCTGACCGGAGACCAACCAAGAGACCGAG 2094
Db ACCGACCGGGCGCGCAGAGAATCGTGAGCCTGACCGGAGACCAACCAAGAGACCGAG 2100
QY GTGAGGGCATCCAGCTGGCCCTGAGGACAGCGGCGAGGGTGAACATCGTGACCGAC 2154
Db GTGAGGGCATCCAGCTGGCCCTGAGGACAGCGGCGAGGGTGAACATCGTGACCGAC 2160
QY AGCCAGTAGCCCTGGGCATCATCAGGCCAGCCGACAGAGCGAGCGAGCTGGTG 2214
Db AGCCAGTAGCCCTGGGCATCATCAGGCCAGCCGACAGAGCGAGCGAGCTGGTG 2220
QY AACCCAGATCATCGACGAGCTGATCAAGAGGAGAGAGTGTACTTGAGTGGTGCCCGCC 2274
Db AACCCAGATCATCGACGAGCTGATCAAGAGGAGAGAGTGTACTTGAGTGGTGCCCGCC 2280
QY CACAAGGGCATCGCGGGCAACAGAGATCGAACAAGCTGGTAGCAAGGGCATCGCGAAG 2334
Db CACAAGGGCATCGCGGGCAACAGAGATCGAACAAGCTGGTAGCAAGGGCATCGCGAAG 2340
QY GTGCTGTTCCTGGAGCGGATCGATCGCGGCGATCGTGATCTACCATGACGACACCTG 2394
Db GTGCTGTTCCTGGAGCGGATCGATCGCGGCGATCGTGATCTACCATGACGACACCTG 2400
QY TACGTGGGCGAGCGGGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCCGGTGA 2454
Db TACGTGGGCGAGCGGGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCCGGTGA 2460
QY TTC 2457
Db TTC 2463
```

```
RESULT 3
AX455914
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 30 from Patent WO020493.
ACCESSION AX455914
VERSION AX455914.1 GI:21714899
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE
1 zur Meede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
polypeptides and uses thereof
JOURNAL Patent: WO 020493-A 30 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
Location/Qualifiers
```

```
source
1. 2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975 (+)"
ORIGIN
Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 3.2e-243;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
QY 1 GTCAGCGCCACCATGCGCGAGCCCATGAGCGGCGCCAGCGCCACCATCTGTATGCAG 60
Db 1 GTCAGCGCCACCATGCGCGAGCCCATGAGCGGCGCCAGCGCCACCATCTGTATGCAG 60
QY 61 GCGAGCAACTTCAAGGGGCCCCAAGGGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
Db 61 GCGAGCAACTTCAAGGGGCCCCAAGGGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGC 120
QY 121 CACATCGCGCCCACTGCGCGCCCGCCCGCAGAGAGGGCTGTGGAAGTGGCGCAAGGAG 180
Db 121 CACATCGCGCCCACTGCGCGCCCGCCCGCAGAGAGGGCTGTGGAAGTGGCGCAAGGAG 180
QY 181 GCGCACCAAGATGAAGGACTGCAACCGAGCGCCAGGCGCAACTTCTTCGCGAGAGACCTGGCC 240
Db 181 GCGCACCAAGATGAAGGACTGCAACCGAGCGCCAGGCGCAACTTCTTCGCGAGAGACCTGGCC 240
QY 241 TTCCCGCAGGGCAAGGCCCGGAGTTCCCGCAGGAGGAGAACCGCGCCCAACAGCCCCCACC 300
Db 241 TTCCCGCAGGGCAAGGCCCGGAGTTCCCGCAGGAGGAGAACCGCGCCCAACAGCCCCCACC 300
QY 301 AGCCCGAGCTGAGGTCGCGCGGAGCAAAACCCCGCAGCGAGCGCGCGCGAGCGCCAG 360
Db 301 AGCCCGAGCTGAGGTCGCGCGGAGCAAAACCCCGCAGCGAGCGCGCGCGAGCGCCAG 360
QY 361 GCGACCTTGAACTTCCCGCAGATCACTGTGGCAGCGCCCTTGTGAGCATCAAGGTG 420
Db 361 GCGACCTTGAACTTCCCGCAGATCACTGTGGCAGCGCCCTTGTGAGCATCAAGGTG 420
QY 421 GCGCGCCAGATCAAGAGGGCCCTGTGTGACACCGCGCGCGAGCACCGTGTGAGGAG 480
Db 421 GCGCGCCAGATCAAGAGGGCCCTGTGTGACACCGCGCGCGAGCACCGTGTGAGGAG 480
QY 481 ATGAGCTTCGCGCGAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540
Db 481 ATGAGCTTCGCGCGAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540
QY 541 GTGCGCCAGTACGACCGATCTGTATCGAGATCTGCGGCAAGAGGCCCATCGGCAACCGTG 600
Db 541 GTGCGCCAGTACGACCGATCTGTATCGAGATCTGCGGCAAGAGGCCCATCGGCAACCGTG 600
QY 601 CTGATCGGCGCCACCCCGTGAAACATCATCGCGCGCAACATGTCACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCGCCACCCCGTGAAACATCATCGCGCGCAACATGTCACCCAGCTGGGCTGC 660
QY 661 ACCCTGAACTTCCCGCAGCCCATCGAGACCGCTGCGCGCGCAAGCTGAAGCCCGCATG 720
Db 661 ACCCTGAACTTCCCGCAGCCCATCGAGACCGCTGCGCGCGCAAGCTGAAGCCCGCATG 720
QY 721 GACGCGCCCAAGTGAAGCTGGGCCCTGACCGGAGAGAGATCAAGGCCCTGACCGCC 780
Db 721 GACGCGCCCAAGTGAAGCTGGGCCCTGACCGGAGAGAGATCAAGGCCCTGACCGCC 780
QY 781 ATCTCGGAGGAGTGGAGAGGGGCAAGATCACCAAGATCGGCGCCCGAGAACCCCTAC 840
Db 781 ATCTCGGAGGAGTGGAGAGGGGCAAGATCACCAAGATCGGCGCCCGAGAACCCCTAC 840
QY 841 AACACCCCGGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 841 AACACCCCGGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCATCCCGCAC 960
Db 901 TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGGAGGTGCACTGGGCATCCCGCAC
```





Query Match		83.0%;	Score 2040;	DB 6;	Length 2300;	
Best Local Similarity		93.6%;	Prod. No. 4.6e-204;			
Matches 2153;		Conservative	0;	Mismatches 135;	Indels 12;	Gaps 2;
170	Y	170	GGGCAAGAGGGCCACACAGATGAAGGACTGACCGAGCGCCAGCCAACTTCTTCCGCG	229		
171	Y	171				
172	Y	172				
173	Y	173				
174	Y	174				
175	Y	175				
176	Y	176				
177	Y	177				
178	Y	178				
179	Y	179				
180	Y	180				
181	Y	181				
182	Y	182				
183	Y	183				
184	Y	184				
185	Y	185				
186	Y	186				
187	Y	187				
188	Y	188				
189	Y	189				
190	Y	190				
191	Y	191				
192	Y	192				
193	Y	193				
194	Y	194				
195	Y	195				
196	Y	196				
197	Y	197				
198	Y	198				
199	Y	199				
200	Y	200				
201	Y	201				
202	Y	202				
203	Y	203				
204	Y	204				
205	Y	205				
206	Y	206				
207	Y	207				
208	Y	208				
209	Y	209				
210	Y	210				
211	Y	211				
212	Y	212				
213	Y	213				
214	Y	214				
215	Y	215				
216	Y	216				
217	Y	217				
218	Y	218				
219	Y	219				
220	Y	220				
221	Y	221				
222	Y	222				
223	Y	223				
224	Y	224				
225	Y	225				
226	Y	226				
227	Y	227				
228	Y	228				
229	Y	229				
230	Y	230				
231	Y	231				
232	Y	232				
233	Y	233				
234	Y	234				
235	Y	235				
236	Y	236				
237	Y	237				
238	Y	238				
239	Y	239				
240	Y	240				
241	Y	241				
242	Y	242				
243	Y	243				
244	Y	244				
245	Y	245				
246	Y	246				
247	Y	247				
248	Y	248				
249	Y	249				
250	Y	250				
251	Y	251				
252	Y	252				
253	Y	253				
254	Y	254				
255	Y	255				
256	Y	256				
257	Y	257				
258	Y	258				
259	Y	259				
260	Y	260				
261	Y	261				
262	Y	262				
263	Y	263				
264	Y	264				
265	Y	265				
266	Y	266				
267	Y	267				
268	Y	268				
269	Y	269				
270	Y	270				
271	Y	271				
272	Y	272				
273	Y	273				
274	Y	274				
275	Y	275				
276	Y	276				
277	Y	277				
278	Y	278				
279	Y	279				
280	Y	280				
281	Y	281				
282	Y	282				
283	Y	283				
284	Y	284				
285	Y	285				
286	Y	286				
287	Y	287				
288	Y	288				
289	Y	289				
290	Y	290				
291	Y	291				
292	Y	292				
293	Y	293				
294	Y	294				
295	Y	295				
296	Y	296				
297	Y	297				
298	Y	298				
299	Y	299				
300	Y	300				
301	Y	301				
302	Y	302				
303	Y	303				
304	Y	304				
305	Y	305				
306	Y	306				
307	Y	307				
308	Y	308				
309	Y	309				
310	Y	310				
311	Y	311				
312	Y	312				
313	Y	313				
314	Y	314				
315	Y	315				
316	Y	316				
317	Y	317				
318	Y	318				
319	Y	319				
320	Y	320				
321	Y	321				
322	Y	322				
323	Y	323				
324	Y	324				
325	Y	325				
326	Y	326				
327	Y	327				
328	Y	328				
329	Y	329				
330	Y	330				
331	Y	331				
332	Y	332				
333	Y	333				
334	Y	334				
335	Y	335				
336	Y	336				
337	Y	337				
338	Y	338				
339	Y	339				
340	Y	340				
341	Y	341				
342	Y	342				
343	Y	343				
344	Y	344				
345	Y	345				
346	Y	346				
347	Y	347				
348	Y	348				
349	Y	349				
350	Y	350				
351	Y	351				
352	Y	352				
353	Y	353				
354	Y	354				
355	Y	355				
356	Y	356				
357	Y	357				
358	Y	358				
359	Y	359				
360	Y	360				
361	Y	361				
362	Y	362				
363	Y	363				
364	Y	364				
365	Y	365				
366	Y	366				
367	Y	367				
368	Y	368				
369	Y	369				
370	Y	370				
371	Y	371				
372	Y	372				
373	Y	373				
374	Y	374				
375	Y	375				
376	Y	376				
377	Y	377				
378	Y	378				
379	Y	379				
380	Y	380				
381	Y	381				
382	Y	382				
383	Y	383				
384	Y	384				
385	Y	385				
386	Y	386				
387	Y	387				
388	Y	388				
389	Y	389				
390	Y	390				
391	Y	391				
392	Y	392				
393	Y	393				
394	Y	394				
395	Y	395				

```
Db 2101 TGGCTTGGTGGCCCGCCCAAGGCGCATCGCGGCAACGAGCAGTGGACAAAGCTGGTGA 2160
Qy 2318 GCAAGGGCATCGCAAGGTGCTGTCTTGGACGGCATCGATGGGGCATCTGATCTACC 2377
Db 2161 GCGCGGATCCGCAAGTGTCTTCTGAACGGCATCGATGGCGCATCTGATCTACC 2220
Qy 2378 AGTACATGACGACCTGTACGTGGGACGCGCGCCCTAGGATCGATTAAAGCTTCCCG 2437
Db 2221 AGTACATGACGACCTGTACGTGGGACGCGCGCCCTAGGATCGATTAAAGCTTCCCG 2280
Qy 2438 GGGCTAGCACCGGTGAATTC 2457
Db 2281 GGGCTAGCACCGGTGAATTC 2300

RESULT 5
CQ870575 2300 bp DNA linear PAT 13-SEP-2004
LOCUS
DEFINITION Sequence 83 from Patent EP1433851.
ACCESSION CQ870575
VERSION CQ870575.1 GI:52000091
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,
AUTHORS Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL Patent: EP 1433851-A 83 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES
source
Location/Qualifiers
1..2300
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
FS(-).protmod.RTopt.YMMW"

ORIGIN

Query Match 83.0%; Score 2040; DB 6; Length 2300;
Best local Similarity 93.6%; Pred. No. 4.6e-204;
Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

Qy 170 GCGCAAGGAGGGCCACCAGATGAGGACTGCACCGAGCGCCAGGCCCACTTCTTCGCG 229
Db 1 GCGCGCGGAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCGCG 60
Qy 230 AGGACCTGGCTTCCCGCAGGCAAGGCCCGCGAGTTCCCGAGGAGCAGAACCGGCCCA 289
Db 61 AGGACCTGGCTTCTGCAAGGCAAGGCCCGCGAGTTGAGGAGGAGCAGACCGGCCCA 120
Qy 290 ACAGCCCAACAGCCGCGAGTGCAGGTGCGCGCG-----ACAACCCCGCAGCGAGG 343
Db 121 ACAGCCCAACCGCGCGAGTGCAGGTGCGCGCGCGAGAACCAACAGCCTGAGCGAGG 180
Qy 344 CCGCGCGGAGCGCCAGGCGACCCCTG-----ACITTCGCCAGATCACCTGTGGCAGC 397
Db 181 CCGCGCGGAGCGCCAGGCGACCGTGTAGCTTCAACTTCCCGCAGATCACCTGTGGCAGC 240
Qy 398 GCCCCTGTGTAGCATCAAGTGGCGGCCAGATCAAGGAGGCGCTCTGACACCGCGC 457
Db 241 GCCCCTGTGTAGCATCAAGTGGCGGCCAGCTCAAGGAGGCGCTCTGACACCGCGC 300
Qy 458 CCGACGACACCGTGTGTGAGGAGATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTGTGTGAGGAGATGAACCTGCGCGCAAGTGAAGCCCAAGATGATCG 360
Qy 518 GCGGCATCGCGGCTTCAATCAAGTGGCGGACGATACGACCGAGATCCTGATCGAGATCTGCG 577
Db 361 GCGGCATCGCGGCTTCAATCAAGTGGCGGACGATACGACCGAGATCCTGCGGATCTGCG 420
```

```
Qy 578 GCAAGAAAGGCCCATCGGCACCGTGTCTGATCGGCCCCACCCCGTGAAACATCATCGGCCGCA 637
Db 421 GCCACAAAGGCATCGGCACCGTGTCTGATCGGCCCCACCCCGTGAAACATCATCGGCCGCA 480
Qy 638 ACATGTGACCCAGCTGGGTGACACCTTGAAATTTCCCATCAGCCCCCATCGAGACCGTGC 697
Db 481 ACCTGTGACCCAGATCGGTGACACCTTGAAATTTCCCATCAGCCCCCATCGAGACCGTGC 540
Qy 698 CCGTGAAGCTGAAGCCCGGCATGACGCGCCCAAGGTGAAAGCTGAGCGCCCTGACCGAGG 757
Db 541 CCGTGAAGCTGAAGCCCGGCATGACGCGCCCAAGGTGAAAGCTGAGCGCCCTGACCGAGG 600
Qy 758 AGAAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCCCTTGATGAGATCTGACCGAGATGGAGAGGAGGCAAGATCACCA 660
Qy 818 AGATCGGCCCGAGAAACCCCTTAACAACCCCGCTGTTCGCCCATCAAGAAAGAGACAGCA 877
Db 661 AGATCGGCCCGAGAAACCCCTTAACAACCCCGCTGTTCGCCCATCAAGAAAGAGACAGCA 720
Qy 878 CCAAGTGGCGCAGCTGTGTGAGCTTCCGCGAGCTGAACAGCGCACCCAGACTTCTGGG 937
Db 721 CCAAGTGGCGCAGCTGTGTGAGCTTCCGCGAGCTGAACAGCGCACCCAGACTTCTGGG 780
Qy 938 AGGTGACAGCTGGGCATCCCCCACCCCGCGCTGAAGAAAGAGAGCGTGAACCGTGC 997
Db 781 AGGTGACAGCTGGGCATCCCCCACCCCGCGCTGAAGAAAGAGAGCGTGAACCGTGC 840
Qy 998 TGAAGTGGCGCAGCTGTGTGAGCTTTCAGCTGCGCTTGAAGAGAGACTTCGCAAGTACACG 1057
Db 841 TGAAGTGGCGCAGCTGTGTGAGCTTTCAGCTGCGCTTGAAGAGAGACTTCGCAAGTACACG 900
Qy 1058 CCTTCAACATCCCGAGCATCAACAGAGACCCCGCGCTTCAGCTGACGATCAACAGTGC 1117
Db 901 CCTTCAACATCCCGAGCATCAACAGAGACCCCGCGCTTCAGCTGACGATCAACAGTGC 960
Qy 1118 TGCCCCAGGCTGGAAGGGCAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1177
Db 961 TGCCCCAGGCTGGAAGGGCAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG 1020
Qy 1178 AGCCCTTCCGCGCGCCCAACCCCGAGATCTGTATACAGGCCCCCTGTGATCGTGGCA 1237
Db 1021 AGCCCTTCCGCAAGCAGAAACCCCGACATCTGTATACAGGCCCCCTGTGATCGTGGCA 1080
Qy 1238 GCGACCTGGAGATCGGCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGC 1297
Db 1081 GCGACCTGGAGATCGGCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGC 1140
Qy 1298 GCTGGGGCTTTCACCAACCCCGCAAGAAAGCAACCAAGAGGAGCCCCCTTCTGCGCCATCG 1357
Db 1141 GCTGGGGCTTTCACCAACCCCGCAAGAAAGCAACCAAGAGGAGCCCCCTTCTGCGCCATCG 1200
Qy 1358 AGTGCACCCCGCAAGTGAACCGTGCAGCCATTCGAGCTGCGCGAGAGAGAGCTGGA 1417
Db 1201 AGTGCACCCCGCAAGTGAACCGTGCAGCCATTCATGCTGCGCGAGAGAGAGCTGGA 1260
Qy 1418 CCGTGAACGACATCCAGAGCTGTGTGGGCAAGCTGAACCTGGCGCAGCAGATCTACCCCG 1477
Db 1261 CCGTGAACGACATCCAGAGCTGTGTGGGCAAGCTGAACCTGGCGCAGCAGATCTACCCCG 1320
Qy 1478 GCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCCAGATCG 1537
Db 1321 GCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCCAGGTGA 1380
Qy 1538 TGCCCTTGCAGGAGGCGGAGCTGAGCTGGCGCGAGAAACCGCGAGATCTGCGCGAGC 1597
Db 1381 TCCCTTGCAGGAGGCGGAGCTGAGCTGGCGCGAGAAACCGCGAGATCTGCGAGGAGC 1440
Qy 1598 CCGTGCAGCGGCTGTACTACGACCCCGAGAGGACCTGTGTGGCGGAGATCCAGAGCAGG 1657
Db 1441 CCGTGCAGCGGCTGTACTACGACCCCGAGAGGACCTGTGTGGCGGAGATCCAGAGCAGG 1500
```

QY	1658	GCCACGACCAAGTGGACCTACCAAGATCTACCAAGAGCCCTTCAAGAACCTTGAAGACCGGCA	1717
Db	1501	GCACGGCCAGTGGACCTACCAAGATCTACCAAGAGCCCTTCAAGAACCTTGAAGACCGGCA	1560
QY	1718	AGTACGCCAAGATGGCCACCGCCACACCAACGACGTGAAGCAGCTGACCCGAGCGCGTGC	1777
Db	1561	AGTACGCCCGCATGCGCGCGCCACACCAACGACGTGAAGCAGCTGACCCGAGCGCGTGC	1620
QY	1778	AGAAAGTCGCCATGAGAGACATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCA	1837
Db	1621	AGAAAGTGAAGACCGAGAGACATCGTATCTGGGGCAAGATCCCAAGTTCGAAGCTGCCCA	1680
QY	1838	TCCAGAAGGAGACCTGGGAGACCTCGTGGACCGACTACTGGCAGGCCACCTGGATCCCGG	1897
Db	1681	TCCAGAAGGAGACCTGGGAGGCTCTGGATGAGTACTTGGCAGGCCACCTTGGATCCCGG	1740
QY	1898	AGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGATCCAGCTGGAGGAAGGAGC	1957
Db	1741	AGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGATCCAGCTGGAGGAAGGAGC	1800
QY	1958	CCATCATCTGGCGCCGAGACCTTCTAGCTGGAGCGCGCCGCAACCGCGAGACCAAGATCG	2017
Db	1801	CCATCTGTGGCGCCGAGACCTTCTAGCTGGAGCGCGCCGCAACCGCGAGACCAAGCTGG	1860
QY	2018	GCAAGGCGGCTACGTGACCGACCGGGCCGGCAGAAGATCGTGAAGCTTGACCCGAGACCA	2077
Db	1861	GCAAGGCGGCTACGTGACCGACCGGGCCGGCAGAAGTGTGAGCATCGCGACACCA	1920
QY	2078	CCAACGAGAAGCCGAGCTGCAAGGCCATCCAGCTGGCCCTGCAGGACACGCGCAGCGAGG	2137
Db	1921	CCAACGAGAAGCCGAGCTGCAAGGCCATCCAGCTGGCCCTGCAGGACACGCGCCTGGAGG	1980
QY	2138	TGAACATGTGACCGACGACGACCTTCTAGCTGGAGCGCGCCGATCATCCAGGCCAGCCCGACAGA	2197
Db	1981	TGAACATGTGACCGACGACGACCTTCTAGCTGGAGCGCGCCGATCATCCAGGCCAGCCCGACAGA	2040
QY	2198	GGGAGCGAGCTGTGTGAACCAAGATCATCGACGACCTGATCAAGAGGAGAGGTGTACC	2257
Db	2041	GGGAGCGAGCTGTGTGAACCAAGATCATCGACGACCTGATCAAGAGGAGAGGTGTACC	2100
QY	2258	TGAGCTGGGTGTCGCCCCCAAGGGCATCGGGGCAACGAGCAGATCGACAAAGCTGGTGA	2317
Db	2101	TGGCTGTGGTGCGCCCCCAAGGGCATCGGGGCAACGAGCAGGTGGCAAGCTGGTGA	2160
QY	2318	GCAAGGGCATCCGCAAGTGTCTTCTGGAGCGCATCGATGGCGGCATCTGTATCTACC	2377
Db	2161	GCGCCGGCATCCGCAAGTGTCTTCTGGAACGGCATCGATGGCGGCATCTGTATCTACC	2220
QY	2378	AGTACATGAGACCTGTACGTGGGCGAGCGGGCCCTAGGATCGATTAAAGCTTCCCG	2437
Db	2221	AGTACATGAGACCTGTACGTGGGCGAGCGGGCCCTAGGATCGATTAAAGCTTCCCG	2280
QY	2438	GGGCTAGCACCGGTGAATTC 2457	
Db	2281	GGGCTAGCACCGGTGAATTC 2300	

RESULT 6  
AR373388  
LOCUS AR373388 2300 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 83 from patent US 6602705.  
ACCESSION AR373388  
VERSION AR373388.1 GI:40075491  
KEYWORDS SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 2300)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
TITLE Expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: US 6602705-A 83 05-AUG-2003;  
FEATURES Location/Qualifiers

source	1. .2300 /organism="unknown" /mol_type="genomic DNA"	
ORIGIN		
Query Match	83.0%; Score 2040; DB 6; Length 2300;	
Best Local Similarity	93.6%; Pred. No. 4.6e-204;	
Matches 2153; Conservative	0; Mismatches 135; Indels 12; Gaps 2;	
QY	170	GCGCAAGGAGGGCCACGATGAAGGACTGCACGAGCGCCAGGCCAACTTCTTCGCGG 229
Db	1	GCGCGCGCAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCGG 60
QY	230	AGGACCTTGGCTTCCCCCAGGCGAGGCCCGGAGTTCCCGAGGAGTCCCGAGCAGAACCCGCGCA 289
Db	61	AGGACCTTGGCTTCTTCGAGGGCAAGGCCCGGAGTTTCAGCAGCAGCAGACCCCGCGCA 120
QY	290	ACAGCCCCACAGCGCGAGCTGCAGGTGCGCGCGG-----ACAACCCCGCAGCGAGG 343
Db	121	ACAGCCCCACCGCGCGAGCTGCAGGTGTCGAGGTGTGGGGCGGAGNAACAACAGCCTGAGCGAGG 180
QY	344	CGGCGCCGAGGCGCAGGGCAACCTG-----AAGTCCCCAGATCAACCTGTGGGAGC 397
Db	181	CGGCGCCGACCGCGCAGGGCAACCGTGAAGCTTCAACTTCCCCAGATCAACCTGTGGGAGC 240
QY	398	GCCCCCTGGTGAAGCATCAAGGTGGCGCGCAGATCAAGGAGGCCCTGCTGGACACCGCGG 457
Db	241	GCCCCCTGGTGAACATCAGGATCGCGCGCAGCTCAAGAGGCGCTGCTCGACACCGCGG 300
QY	458	CCGACGACACCGTGTGGAGGAGATGAGCTGCGCGCAAGTGGAGGCCCAAGATGATCG 517
Db	301	CCGACGACACCGTGTGGAGGAGATGAACTGCGCGCAAGTGGAGGCCCAAGATGATCG 360
QY	518	GCGGCATCGCGCGCTTCAAGGTGCGCCAGTAGACACAGATCTCTGATCGAGATCTGGG 577
Db	361	GCGGCATCGCGCGCTTCAAGGTGCGCGCAGTAGACACAGATCTCTGATCGAGATCTGGG 420
QY	578	GCAGAGGCCATCGGCACCGTGTGATCGCGCCCGCCACCGCTGAACATCATCGGCCCA 637
Db	421	GCCAGAGGCCATCGGCACCGTGTGATCGCGCCCGCCACCGCTGAACATCATCGGCCCA 480
QY	638	ACATGCTGACCCAGCTGGGCTGCACCTTGAATTTCCCATCAGCCCCCATCGAGACCGTGC 697
Db	481	ACCTGCTGACCCAGATCGGCTGCACCTTGAATTTCCCATCAGCCCCCATCGAGACCGTGC 540
QY	698	CCGTGAAGCTGAAGCCCGGCATGGAAGCCCGCAAGGTGAAGCAGTGGGCCCTTGAACGAGG 757
Db	541	CCGTGAAGCTGAAGCCCGGCATGGAAGCCCGCCCAAGGTCAAGCAGTGGGCCCTTGAACGAGG 600
QY	758	AGAGATCAAGCCCTGACCGCCATCTCGAGGAGATGGAGAGGAGGCGCAAGATCACCA 817
Db	601	AGAGATCAAGCCCTTGTGGAGATCTGACCGAGATGGAGAGGAGGCGCAAGATCACCA 660
QY	818	AGATCGGCCCGAGAACCCCTACAAACCCCGGTGTTCCCATCAAGAGAGGAGCAGCA 877
Db	661	AGATCGGCCCGAGAACCCCTTACAAACCCCGGTGTTCCCATCAAGAGAGGAGCAGCA 720
QY	878	CAAAGTGGCGAAGCTGTGTGACTTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 937
Db	721	CAAAGTGGCGAAGCTGTGTGACTTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG 780
QY	938	AGGTGACGTGGGCATCCCCCAGCCCGGCTTGAAGAGAGAGAGCGTGAACCGTGC 997
Db	781	AGGTGACGTGGGCATCCCCCAGCCCGGCTTGAAGAGAGAGAGCGTGAACCGTGC 840
QY	998	TGGACGTGGGCGAGCGCTACTTTCAGCGTGGCCCCCTGGACGAGGACTTCCGCAAGTACACG 1057
Db	841	TGGACGTGGGCGAGCGCTACTTTCAGCGTGGCCCCCTGGACAGGACTTCCGCAAGTACACG 900
QY	1058	CCTTCACCATCCCGAGCATCAACAACGAGACCCCGGCTCGGCTACCGATCAACGAGTGC 1117
Db	901	CCTTCACCATCCCGAGCATCAACAACGAGACCCCGGCTCGGCTACCGATCAACGAGTGC 960

Qy	1118	TGCCCCAGGGCTGGAAAGCGCAGCCCGACGCATCTTTCCAGAGCAGCATGACCAAGATCTCG	1177
Db	961	TGCCCCAGGGCTGGAAAGCGCAGCCCGCATCTTTCAGAGCAGCATGACCAAGATCTCTGG	1020
Qy	1178	AGCCCTTCCGCGCCGCAACCCCAGATCGTGATCTACCAGGCCCCCTCTACTGTGGGCA	1237
Db	1021	AGCCCTTCCGCGAGAGAACCCCAGCATCTGTATCTACCAGGCCCCCTCTACTGTGGGCA	1080
Qy	1238	GCGACCTGGAGATCGGCGCAGCACCGCGCCCAAGATCGAGAGGCTGCGCAAGCATCTGCTGC	1297
Db	1081	GCGACCTGGAGATCGGCGCAGCACCGCCACCAAGATCGAGAGGCTGCGCCAGCATCTGCTGC	1140
Qy	1298	GCTGGGGTTCCACACCCCGCAAGAAGACACCAAGAGGAGGCCCTTCTGTGCCAATCG	1357
Db	1141	GCTGGGGTTTCCACACCCCGCAAGAAGACACCAAGAGGAGGCCCTTCTGTGCCAATCG	1200
Qy	1358	AGCTGCACCCCGACAAGTGGACCGTGAGCCCATCGAGCTGCCCGAAGAGAGGAGCTGGA	1417
Db	1201	AGCTGCACCCCGACAAGTGGACCGTGAGCCCATCATGCTGCCCGAAGAGAGGAGCTGGA	1360
Qy	1418	CCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGGCAGCGCAGATCTACCCCG	1477
Db	1261	CCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGAGCCAGATCTACCGCG	1320
Qy	1478	GCATCAAAGTGGCCAGCTGTGCAAGCTGTGCGCGCGCCAAGGCCCTTCAACCGACATCG	1537
Db	1321	GCATCAAAGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCCAAGGCCCTTCAACCGAGTGA	1380
Qy	1538	TGCCCTTGAACGAGGCGGAGCTGGAGCTGGCCGAGAACCGGAGATCTCTGGCGAGC	1597
Db	1381	TCCCCCTTGAACGAGGCGGAGCTGGAGCTGGCCGAGAACCGGAGATCTCTGAAGGAGC	1440
Qy	1598	CCGTGCACGGCGTGTACTACGACCCAGCAAGGACCTGTGTGGCGAGATCCAGAAAGCAGG	1657
Db	1441	CCGTGCACGGCGTGTACTACGACCCAGCAAGGACCTGTGTGGCGAGATCCAGAAAGCAGG	1500
Qy	1658	GCCACGACAGTGAACCTTACAGATCTACAGGAGCCCTTCAAGAACTTCAAGACCGGCA	1717
Db	1501	GCCAGGCGCAGTGGACCTTACAGATCTACAGGAGCCCTTCAAGAACTTCAAGACCGGCA	1560
Qy	1718	AGTAGGCCAAGATGCGCACCGCCACACCAACGACGTGAAGCAGCTGACCGAGCCGTGC	1777
Db	1561	AGTAGGCCCGCATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGCCGTGC	1620
Qy	1778	AGAAATGCGCATGEGAGAGATCTGTATCTGGGGCAAGACCCCAAGTCTCCGCTGCCCCA	1837
Db	1621	AGAAAGTGAAGCAGAGAGATCTGTATCTGGGGCAAGATCCCAAGTTCAAGCTGCCCCA	1680
Qy	1838	TCCAGAAAGGAGACTGGGAGACTGTGTGACCGCATCTGGCAGGCCACCTGGATCCCCG	1897
Db	1681	TCCAGAAAGGAGACTGGGAGGCCCTGTGTGATGAGTACTGGCAGGCCACCTGGATCCCCG	1740
Qy	1998	AGTGGAGTTCTGTGAACACCCGCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAGAGAGC	1957
Db	1741	AGTGGAGTTCTGTGAACACCCGCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAGAGAGC	1800
Qy	1958	CCATCATCGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCAAGATCG	2017
Db	1801	CCATCTGTGGCGCGAGACCTTCTACGTGGACGCGCGCCCAACCGCGAGACCAAGCTGG	1860
Qy	2018	GCAAGGCCGCTACTGTACCGACCGAGCGGGCGGCGAGAGATCTGTAGCTTCAACCGAGACCA	2077
Db	1861	GCAAGGCCGCTACTGTACCGACCGAGCGGGCGGCGAGAGAGTGTGTAGCTTCAACCGACCA	1920
Qy	2078	CCAAACGAGAGACCGAGCTGAGGCCATTCAGCTGTGGCCCTTGCAGGACAGCGGAGCGAGG	2137
Db	1921	CCAAACGAGAGACCGAGCTGAGGCCCATTCACCTTGGCCCTTGCAGGACAGCGGCTTGGAGG	1980
Qy	2138	TGAACATCTGTACCGACAGCAGTACGCCCTTGGGCGATCATCCAGGCCACCGCAAGA	2197
Db	1981	TGAACATCTGTACCGACAGCAGTACGCCCTTGGGCGATCATCCAGGCCACCGCAAGA	2040
Qy	2198	GCGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGTATCAAGAGGAGAGAGTGTATCC	2257

Db	2041	GCAGAGCGAGCTGGTGGAGCCAGATCATCGAGCAGCTGATCAAGAGAGGAAGGTGTACC	2100
Qy	2258	TGAGCTGGGTGCGCGCCCAAGGGCATCGCGCGACAGAGAGATCGACAAGCTGGTGA	2317
Db	2101	TGGCTGGGTGCGCGCCCAAGGGCATCGCGCGACAGAGAGTGGACAAGCTGGTGA	2160
Qy	2318	GCAAGGGCATCCGCAAGGTGCTGTTCTTGACGGGCATCGATGGCGGCATCGTGTATCTACC	2377
Db	2161	GCGCCGCATCCGCAAGGTGCTGTTCTGAACGGGCATCGATGGCGGCATCGTGTATCTACC	2220
Qy	2378	AGTACATGGAGACCTGTAGTGGCAGCGCGGCCCTTAGAGTCGATTAAGAGTTCCCG	2437
Db	2221	AGTACATGGAGACCTGTAGTGGCAGCGCGGCCCTTAGAGTCGATTAAGAGTTCCCG	2280
Qy	2438	GGGCTAGCACCGGTGAATTC	2457
Db	2281	GGGCTAGCACCGGTGAATTC	2300
RESULT 7			
BD263704			
LOCUS	BD263704	2306 bp DNA linear	PAT 17-JUL-2003
DEFINITION	Improved expression of HIV polypeptides and production of virus-like particles.		
ACCESSION	BD263704		
VERSION	BD263704.1	GI:33073472	
KEYWORDS	JP 2002533124-A/71.		
SOURCE	synthetic construct		
ORGANISM	other sequences; artificial sequences.		
REFERENCE	1 (bases 1 to 2306)		
AUTHORS	Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.		
TITLE	Improved expression of HIV polypeptides and production of virus-like particles		
JOURNAL	Patent: JP 2002533124-A 71	08-OCT-2002;	
COMMENT	CHIRON CORP		
	OS Artificial Sequence		
	PN JP 2002533124-A/71		
	PD 08-OCT-2002		
	PF 30-DEC-1999 JP 2000591193		
	PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471	PI	
	SUSAN BARNETT,JAN ZUR MEGEDE,INDRESH SRIVASTAVA,YING LIAN, PI		
	KARIN HARTOG,		
	PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER	PC	
	C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61F31/18,A61F37/02, PC		
	C12N5/10,		
	PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC		
	Description of Artificial Sequence: FS(-).protmod.RTopt.YM FH	Key	
	Location/Qualifiers		
FT	source	1..2306	
FT	Location/Qualifiers		
	Location/Qualifiers		
	1..2306		
	/organism="synthetic construct"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:32630"		
ORIGIN			
Query Match	82.2%	Score 2019.2;	DB 6; Length 2306;
Best Local Similarity	93.2%;	pred. No. 6.8e-202;	
Matches 2150;	Conservative	0; Mismatches 138;	Indels 18; Gaps 3;
Qy	170	GCGCCAAGGAGGCGCACCATGATGAAGACTGCGACCGAGCGCCAGGCCAATCTTCTCGCG	229
Db	1	GCGCCCGGAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG	60
Qy	230	AGGACCTGGCTTCCCCCAGGCAAGCGCCGCGAGTTCCCGAGGAGCAGCAACCGGCCCA	289
Db	61	AGGACCTGGCTTCTCTGCGGGAAGGCCCGCGAGTTACAGCGGAGCAGACCCCGGCCA	120
Qy	290	ACAGCCCCACGAGCGCGAGCTGCAGGTGCGCGCGC-----ACAAACCCCGCAGCGAGG	343



RESULT 8  
 CQ870574 2306 bp DNA linear PAT 13-SEP-2004  
 LOCUS  
 DEFINITION Sequence 82 from Patent EP1433851.  
 ACCESSION CQ870574  
 VERSION CQ870574.1 GI:52000090  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 synthetic construct  
 other sequences; artificial sequences.  
 1  
 REFERENCE  
 AUTHORS Barnett,S., Zurnegede,J., Srivastava,I., Lian,Y., Hartog,K.,  
 Liu,H., Greer,C., Selby,M. and Walker,C.  
 TITLE Improved expression of HIV polypeptides and production of  
 virus-like particles  
 JOURNAL Patent: EP 1433851-A 82 30-JUN-2004;  
 CHIRON CORPORATION (US)  
 FEATURES  
 source  
 1.2306  
 Location/Qualifiers  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="Description of Artificial Sequence:  
 FS(-).protmod.RTopt.YM"  
 ORIGIN  
 Query Match 82.2%; Score 2019.2; DB 6; Length 2306;  
 Best Local Similarity 93.2%; Pred. No. 6.8e-202;  
 Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
 QY 170 GCGCAAGGAGGCGCCACAGATGAAGACTGCACCGAGCGCGAGCCCAACTTCTTCGCG 229  
 DB 1 GCGCGCGGAGGACACCAATGAAGATTGCATGAGAGCAGGCTAATTTCTTCGCG 60  
 QY 230 AGGACCTGGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGAGCAGACACCGGCCA 289  
 DB 61 AGGACCTGGGCTTCTGCGGCGAAGCCCGCGAGTTTCAGCAGCGAGCAGACCGGCCA 120  
 QY 290 ACAGCCCCACGCGCGAGCTGAGGTGGCGGCGG-----ACAAACCCCGGAGCGAGG 343  
 DB 121 ACAGCCCCACCGCGCGAGCTGAGGTGGCGGCGGAGAAACAGCGCTGAGCGAGG 180  
 QY 344 CCGCGCGCGAGCGCGAGCGACCGT-----AACTTCCCGCAGATCACTCTGTGGCAGC 397  
 DB 181 CCGCGCGCGAGCGCGAGCGACCGTGAAGTTCACCTTCCCGCAGATCACTCTGTGGCAGC 240  
 QY 398 GCGCGCGCGAGCGCGAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 457  
 DB 241 GCGCGCGCGAGCGCGAGCG 300  
 QY 458 CCGAGGACCGCTGCTGAGGAGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 517  
 DB 301 CCGAGGACCGCTGCTGAGGAGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360  
 QY 518 GCGCGATCGCGCGCTTCATCAAGTGGCGCGAGTACGACCGAGTCCGATCGAGATCTGCG 577  
 DB 361 GCGCGATCGCGCGCTTCATCAAGTGGCGCGAGTACGACCGAGTCCGATCGAGATCTGCG 420  
 QY 578 GCAAGAGGCGCATCGGCG 637  
 DB 421 GCAAGAGGCGCATCGGCG 480  
 QY 638 ACATGCTGACCGAGCTGGGTGACACCTTGAATTTCCCGATCAGCCCCATCGAGACCGTGC 697  
 DB 481 ACCTGCTGACCGAGTGGGTGACACCTTGAATTTCCCGATCAGCCCCATCGAGACCGTGC 540  
 QY 698 CCGTGAAGCTGAGCG 757  
 DB 541 CCGTGAAGCTGAGCG 600  
 QY 758 AGAAGATCAAGGCGCGCTGACCGCGCATCTGCGAGGAGATGGAGAGGCGGCAAGATCACCA 817

DB 601 AGAAGATCAAGGCGCGCTGCTGGAGATCTGCAACCGAGATGGAGAGGCGCAAGATCAGCA 660  
 QY 818 AGATCGCGCGCGGAAACCCCTTACAAACCCCGCTGTTCCGCCATCAAGAGAGGAGCAGCA 877  
 DB 661 AGATCGCGCGCGGAAACCCCTTACAAACCCCGCTGTTCCGCCATCAAGAGAGGAGCAGCA 720  
 QY 878 CCAAGTGGCGCAAGCTGCTGGAGCTTCCGCGAGCTGAAACAGCGCACCCAGAGCTTCTGGG 937  
 DB 721 CCAAGTGGCGCAAGCTGCTGGAGCTTCCGCGAGCTGAAACAGCGCACCCAGAGCTTCTGGG 780  
 QY 938 AGTGAAGCTGGGATCTCCCGCAACCCCGCGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 997  
 DB 781 AGTGAAGCTGGGATCTCCCGCAACCCCGCGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 840  
 QY 998 TGAAGCTGGGCGAGCGCTTACTTTCAGCGTGGCGCGAGCTTCCGCAAGTACACCG 1057  
 DB 841 TGAAGCTGGGCGAGCGCTTACTTTCAGCGTGGCGCGAGCTTCCGCAAGTACACCG 900  
 QY 1058 CTTTACCATCCCGAGCATCAACAAACAGAGACCCCGCGCATCCGCTACCAAGTACCAACGTGC 1117  
 DB 901 CTTTACCATCCCGAGCATCAACAAACAGAGACCCCGCGCATCCGCTACCAAGTACCAACGTGC 960  
 QY 1118 TGCCCCAGGCTGGAAGGCGAGCGCGCGCATCTTTCAGAGAGAGAGATGACCAAGATCTCTGG 1177  
 DB 961 TGCCCCAGGCTGGAAGGCGAGCGCGCGCATCTTTCAGAGAGAGAGATGACCAAGATCTCTGG 1020  
 QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATCTGATCTACCAAGGCGCGCGCTTACCTGGGCA 1237  
 DB 1021 AGCCCTTCCGCGCGCGCAACCCCGAGATCTGATCTACCAAGGCGCGCGCTTACCTGGGCA 1080  
 QY 1238 GCGACCTGGAGATCGGCGAGCAGCGCGCAAGATCGAGGAGCTCGCAAGCAGCTCTGTCG 1297  
 DB 1081 GCGACCTGGAGATCGGCGAGCAGCGCGCAAGATCGAGGAGCTCGCAAGCAGCTCTGTCG 1140  
 QY 1298 GCTGGGCTTTCACACCCCGCAAGAGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1355  
 DB 1141 GCTGGGCTTTCACACCCCGCAAGAGAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 QY 1356 ----CGAGCTGCAACCCCGCAAGTGGAGCTGAGCGCGCGCATCGAGCTGCGCGAGAGAGAG 1411  
 DB 1201 GCTACGAGCTGCAACCCCGCAAGTGGAGCTGAGCGCGCGCATCGATCTGCTGCGGAGAGAG 1260  
 QY 1412 GCTGGAGCCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACTGGCGCGAGCAGATCT 1471  
 DB 1261 GCTGGAGCCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACTGGCGCGAGCAGATCT 1320  
 QY 1472 ACCCGGCGATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCGTGAACCG 1531  
 DB 1321 ACGCGGCGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCGTGAACCG 1380  
 QY 1532 ACATCGTGGCGTGAACCGAGGCGCGAGCTGGAGCTGGCGCGAGAGAGAGAGAGAGAGAGAG 1591  
 DB 1381 AGTGAATCCCGCTGACCGAGGAGCGAGCTGGAGCTGGCGCGAGAGAGAGAGAGAGAGAGAG 1440  
 QY 1592 GCGAGCCCGTGAACCGCGCTGTACTACGACCCCGAGAGAGAGCTGGTGGCGCGAGATCAGAG 1651  
 DB 1441 AGAGCCCGTGAACCGAGTGTACTACGACCCCGAGAGAGAGCTGGTGGCGCGAGATCAGAG 1500  
 QY 1652 AGAGGCGCGAG 1711  
 DB 1501 AGCAGGCGCGAGGCGAGTGGAGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
 QY 1712 CCGCGAG 1771  
 DB 1561 CCGCGAG 1620  
 QY 1772 CCGTGAAG 1831  
 DB 1621 CCGTGAAG 1680  
 QY 1832 TGCCCATCCAG 1891



Db 1681 TGCCCATCCAGAAAGAGACTCTGGAGGCTGTGGATGGAGTACTGGCAGGCGCACTGGGA 1740  
Qy 1892 TCCCGAGTGGAGTTCGTGAACACCCCGCCCTCGTGAAGCTGTGGTACAGCTGGAGA 1951  
Db 1741 TCCCGAGTGGAGTTCGTGAACACCCCGCCCTCGTGAAGCTGTGGTACAGCTGGAGA 1800  
Qy 1952 AGGAGCCCATCATCGCGGCCGAGAGACTTCTAGCTGGACGGCGCCCAACCGCGAGACCA 2011  
Db 1801 AGGAGCCCATCGTGGCGCCGAGAGCTTCTAGCTGGACGGCGCCCAACCGCGAGACCA 1860  
Qy 2012 AGATCGGCAAGCCCGCTGCTGTCGACCGAGACCGGGCCGGCAGAGATCTGTGAGCTTGACCG 2071  
Db 1861 AGCTGGGCAAGCCCGCTGCTGTCGACCGAGACCGGGCCGGCAGAGATCTGTGAGCTTGACCG 1920  
Qy 2072 AGACCACCAACAGAAAGCCGAGCTGTCAGGCGCATCCAGCTGGCCCTGTCAGGACAGCGGCA 2131  
Db 1921 ACACCACCAACAGAAAGCCGAGCTGTCAGGCGCATCCAGCTGGCCCTGTCAGGACAGCGGCC 1980  
Qy 2132 GCGAGTGAACATCGTGAACCGACAGCCAGTACGCCCTGGGCATCATCAGGCGCCAGCGCCG 2191  
Db 1981 TGGAGTGAACATCGTGAACCGACAGCCAGTACGCCCTGGGCATCATCAGGCGCCAGCGCCG 2040  
Qy 2192 ACAAGAGCGAGAGCGAGCTGTGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAAAG 2251  
Db 2041 ACAAGAGCGAGAGCGAGCTGTGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAAAG 2100  
Qy 2252 TGTACTGAGCTGGGTGCGGCCCAAGGGGATCGCGGCATCGATCGCGGCATCTGTGA 2371  
Db 2101 TGTACTGAGCTGGGTGCGGCCCAAGGGGATCGCGGCATCGATCGCGGCATCTGTGA 2220  
Qy 2372 TGTACTGAGCTGGGTGCGGCCCAAGGGGATCGCGGCATCGATCGCGGCATCTGTGA 2371  
Db 2161 TGTACTGAGCTGGGTGCGGCCCAAGGGGATCGCGGCATCGATCGCGGCATCTGTGA 2220  
Qy 2372 TGTACTGAGCTGGGTGCGGCCCAAGGGGATCGCGGCATCGATCGCGGCATCTGTGA 2371  
Db 2221 TGTACTGAGCTGGGTGCGGCCCAAGGGGATCGCGGCATCGATCGCGGCATCTGTGA 2220  
Qy 2432 TTCCCGGGGCTAGCACCGGTGAATTC 2457  
Db 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 9  
AR373387  
LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 82 from patent US 6602705.  
ACCESSION AR373387  
VERSION AR373387.1 GI:40075490  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2306)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
TITLE Expression of HIV polypeptides and production of virus-like  
PARTICLES  
JOURNAL Patent: US 6602705-A 82 05-AUG-2003;  
FEATURES  
Location/Qualifiers  
1..2306 /organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 82.2%; Score 2019.2; DB 6; Length 2306;  
Best Local Similarity 93.2%; Pred. No. 6.8e-202;  
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
Qy 170 GCGCAAGGAGGCGCCACAGATGAAGACTGCACCGAGCGCCAGGCGCACTTCTCCGG 229  
Db 1 GCGGCGGAGGAGCACCAATATGAAGATTGCACTGAGAGAGCTGAATTTCTTCCGG 60  
Qy 230 AGGACTGCGCTTCCCGGAGGCAAGGCGCGGAGTTCCCGAGGAGCAACCGCGCCA 289

Db 61 AGGACTGCGCTTCTCTGCAAGGCGCAAGGCGCGAGTTCAAGCAGAGCAGACCCCGCCA 120  
Qy 290 ACAGCCCCACAGCGCGAGCTGCAGGTGCGGGG-----ACAAACCCCGCAGCGAGG 343  
Db 121 ACAGCCCCACAGCGCGAGCTGCAGGTGCGGGGCGGAGAACAAACAGCTGAGCGAGG 180  
Qy 344 CCGGCGCGAGCGCGCAGGCGCACCCCTG-----AACTTCCCCAGATCACCCCTGTGGCAGC 397  
Db 191 CCGGCGCGAGCGCGCAGGCGCACCGTGAGCTTCACTTCCCCAGATCACCCCTGTGGCAGC 240  
Qy 398 GCGCCCTGTGAGCATCAAGGTGGCGCGCAGATCAAGGAGGCGCTGTGTGACACCGCGC 457  
Db 241 GCGCCCTGTGAGCATCAAGGTGGCGCGCAGTCAAGGAGGCGCTGTGTGACACCGCGC 300  
Qy 458 CCGAGCACCGTGTGTGAGGAGATGAGCTTCCCGGCAAGTGGGAAGCCCAAGATGATCG 517  
Db 301 CCGAGCACCGTGTGTGAGGAGATGAACTTCCCGGCAAGTGGGAAGCCCAAGATGATCG 360  
Qy 518 GCGGCATCGCGCGCTTTCATCAAGGTGGCGCGCAGTACGACAGATCCTGATCCGAGATCTGGC 577  
Db 361 GCGGATCGGGGCTTTCATCAAGGTGGCGCGCAGTACGACAGATCCCCGTGGAGATCTGGC 420  
Qy 578 GCAAGAGGCCATCGGCAACCGTGTGTGATCGGCCCCCACCCCGTGAACATCATCGCGCGCA 637  
Db 421 GCCAAGGCCCATCGGCAACCGTGTGTGATCGGCCCCCACCCCGTGAACATCATCGCGCGCA 480  
Qy 638 ACATGCTGACCGAGCTGGGCTGCACCTCGAATCTTCCCATCAGCCCCCATCGAGACCGTGC 697  
Db 481 ACCTGCTGACCGAGATCGGCTGCACCTCGAATCTTCCCATCAGCCCCCATCGAGACCGTGC 540  
Qy 698 CCGTGAAGCTGAAGCCCGGCGCATGGACGCGCCCAAGGTGAAGCAGTGGGCCCTGACCGAGG 757  
Db 541 CCGTGAAGCTGAAGCCCGGCGCATGGACGCGCCCAAGGTGAAGCAGTGGGCCCTGACCGAGG 600  
Qy 758 AGAAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGGAGAAGAGGCGCAAGATCAACA 817  
Db 601 AGAAGATCAAGGCCCTGATGGAGATCTGCAACCGAGATGGAGAAGAGGCGCAAGATCAGCA 660  
Qy 818 AGATCGGCGCGAGAACCCCTTACAACACCCCGGTGTTCCGCATCAAGAGAGGAGCAGCA 877  
Db 661 AGATCGGCGCGAGAACCCCTTACAACACCCCGGTGTTCCGCATCAAGAGAGGAGCAGCA 720  
Qy 878 CCAAGTGGCGCAAGCTGTGTGACTTCCCGCAGCTGAACAAGCGCACCCAGAGCTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGTGTGACTTCCCGCAGCTGAACAAGCGCACCCAGAGCTTCTGGG 780  
Qy 938 AGGTGAGCTGGGCATCCCCCAACCGCGCGCTGAAGAAGAGAGCGGTGACCGTGC 997  
Db 781 AGGTGAGCTGGGCATCCCCCAACCGCGCGCTGAAGAAGAGAGCGGTGACCGTGC 840  
Qy 998 TGGAGCTGGGCGACGCTTACTTTCAGCGTGGCCCTGGAGCAGGACTTCCGCAAGTACACCG 1057  
Db 841 TGGAGCTGGGCGACGCTTACTTTCAGCGTGGCCCTGGAGCAGGACTTCCGCAAGTACACCG 900  
Qy 1058 CTTTCACCATCCCGAGCATCAACACAGAGACCCCGGCATCCGCTACCAAGTACCAAGTGC 1117  
Db 901 CTTTCACCATCCCGAGCATCAACACAGAGACCCCGGCATCCGCTACCAAGTACCAAGTGC 960  
Qy 1118 TGCCCCAGGCGTGGAAAGGCGAGCCCCCAGCATTTTCCAGAGCAGCATGACCAAGATCTCTGG 1177  
Db 961 TGCCCCAGGCGTGGAAAGGCGAGCCCCCAGCATTTTCCAGAGCAGCATGACCAAGATCTCTGG 1020  
Qy 1178 AGCCCTTCGCGCGCGCAACCCCGAGATCGTGATCTACAGGCCCGCTGTACGTTGGCA 1237  
Db 1021 AGCCCTTCGCGCAAGAGAACCCCGCATCTGTGATCTACAGGCCCGCTGTACGTTGGCA 1080  
Qy 1238 GCGACTGGAGATCGGCGCAGACACCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGC 1297  
Db 1081 GCGACTGGAGATCGGCGCAGACACCGCAAGATCGAGAGCTGCGCGAGCACTGCTGC 1140  
Qy 1298 GCTGGGCTTCAACACCCCGCAAGAGACACCAAGAGGAGCCCCCTTCTCTGCCCAT-- 1355









Db	1861	AGACCAAGCTGGGCAAGCGCGCTACTGTCACCGACCGGGCGGCAGAAAGTGGTGAGCA	1920	Qy	344	CGGCGCCGAGCGCGCAGGCGACCCCTG-----AACTTCCCCAGATCACCTGTGGCAGC	397
Qy	2066	TCACCGAGACCAACACAGAGACCGAGCTGCAGGCCATCAGCTGGCCCTGCAGGACA	2125	Db	191		
Db	1921	TCGCCGACACCAACACAGAGACCGAGCTGCAGGCCATCACCCTGGCCCTGCAGGACA	1980	Qy	398	CGCCCTTGGTGAAGCATCAAGGTGGCGCGCCAGATCAAGGAGGCCCTGTGTGACACACGGCG	457
Qy	2126	GGCGCAGCGAGGTGAACATCGTGAACCGACGACGAGTACGCCCTGGGCATCATCCAGGGCCC	2185	Db	241	CGCCCTTGGTGAACCATCAGGATCGCGCGCCAGCTCAAGGAGCGCTGTCTGCACACCGCGC	300
Db	1981	GGCGCTGGAGGTGAACATCGTGAACCGACGACGAGTACGCCCTGGGCATCATCCAGGGCCC	2040	Qy	458	CGGACGACACCGTCTGGAGGAGATGAGCTTCCCGGCAAGTGGAGGCCCAAGATGATCG	517
Qy	2186	AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACACAGATCATGAGCAGCTGATCAAGAGG	2245	Db	301	CGGACGACACCGTCTGGAGGAGATGAACCTGCGCGGCAAGTGGAGGCCCAAGATGATCG	360
Db	2041	AGCCCGACAAGAGCGAGAGCGAGCTGGTGAACCGACGAGTACGCCCTGGGCATCATCCAGGGCCC	2100	Qy	518	CGGCGATCGCGCGCTTCAATCAAGGTGGCGCAGTACGACAGATCTGTGATCGAGATCTGCG	577
Qy	2246	AGAAGGTGTACTGAGCTGGGTGGCGGCCCAAGGGCATCGCGCGCAACGACGATCG	2305	Db	361	CGGCGATCGCGGGCTTCAATCAAGGTGGCGCAGTACGACAGATCTGTGATCGAGATCTGCG	420
Db	2101	AGAAGGTGTACTGCGCTGGGTGGCGGCCCAAGGGCATCGCGCGCAACGACGAGGTGG	2160	Qy	578	GCAAGAGGCGCATCGGCGACCGTGTGATCGGCGGCCACCCCGTGAAATCATTCATCGGCGCGCA	637
Qy	2306	ACAAAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGTCTGAGCGGCATCGATGGCGGCA	2365	Db	421	GCCACAAAGGCCATCGGCGACCGTGTGTGGGGCGGCCACCCCGTGAAATCATTCATCGGCGCGCA	480
Db	2161	ACAAAGCTGTGAGCGCGCGCATCCGCAAGGTGCTGTCTGAGCGGCATCGATGGCGGCA	2220	Qy	638	ACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGC	697
Qy	2366	TGCTGATCTACAGTACATGAGACGACCTGTACGTGGCGAGCGCGCCCTAGGATCGATT	2425	Db	481	ACCTGCTGACCCAGATCGGCTGCACCTTGAATTTCCCATCAGCCCATCGAGACCGTGC	540
Db	2221	TGCTGATCTACAGTACATGAGACGACCTGTACGTGGCGAGCGCGCCCTAGGATCGATT	2280	Qy	698	CGGTGAAGCTGAAGCCCGGCATGGACGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGGAGG	757
Qy	2426	AAAAGCTTCCCGGGCTAGCACCGGTGAATTC	2457	Db	541	CGGTGAAGCTGAAGCCCGGGATGGACGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGGAGG	600
Db	2281	AAAAGCTTCCCGGGCTAGCACCGGTGAATTC	2312	Qy	758	AGAAAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAGAAAGAGGGCAAGATCACCA	817
RESULT 12				Db	601	AGAAAGATCAAGGCGCTTGGTGGAGATCTGCACCGAGATGGAGAAAGAGGGCGCAAGATCAGCA	660
LOCUS	Q0870576	2312 bp	DNA	Qy	818	AGATCGGCGCGAGAAACCCCTACACACCCCGGTGTTCCCATCAAGAGAGAGGACAGCA	877
DEFINITION	Sequence 84 from Patent EP1433851.		linear	Db	661	AGATCGGCGCGAGAAACCCCTACACACCCCGGTGTTCCCATCAAGAGAGAGGACAGCA	720
ACCESSION	Q0870576			Qy	878	CAAAGTGGCGCAAGCTGGTGGACTTCCCGGAGCTGAACAAGCGCACCCAGGACTTCTGGG	937
VERSION	Q0870576.1	GI:52000092		Db	721	CAAAGTGGCGCAAGCTGGTGGACTTCCCGGAGCTGAACAAGCGCACCCAGGACTTCTGGG	780
KEYWORDS	synthetic construct			Qy	938	AGGTGAGCTGGGCATTCGCCACCCCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGC	997
ORGANISM	other sequences; artificial sequences.			Db	781	AGGTGAGCTGGGCATTCGCCACCCCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGC	840
REFERENCE	1			Qy	998	TGGAGCTGGGCGAGCGCTTCTGAGCGTCCCTGGAGAGGAGCTTCCGCGAAGTACACCG	1057
AUTHORS	Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.			Db	841	TGGAGCTGGGCGAGCGCTTCTGAGCGTCCCTGGAGAGGAGCTTCCGCGAAGTACACCG	900
TITLE	Improved expression of HIV polypeptides and production of virus-like particles			Qy	1058	CCTTCACCATCCCGAGCATCAACAACGAGACCCCGGCATCCGCTACAGTACACAGTGC	1117
JOURNAL	Patent: EP 1433851-A 84 30-JUN-2004; CHIRON CORPORATION (US)			Db	901	CCTTCACCATCCCGAGCATCAACAACGAGACCCCGGCATCCGCTACAGTACACAGTGC	960
FEATURES	Location/Qualifiers			Qy	1118	TGCCCGAGGTGAAGGCGAGCCCGACATCTTCCAGAGCAGCATGACCAAGATCTCTGG	1177
source	1..2312 /organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="Description of Artificial Sequence: FS(-).protmod.Rtopt(+)"			Db	961	TGCCCGAGGTGAAGGCGAGCCCGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG	1020
ORIGIN				Qy	1178	AGCCCTTCCGCGCGCAACCCCGAGATCGTGTACTACCA-----GGCCCCCTGTAGC	1231
Query Match	81.3%;	Score 1998.4;	DB 6;	Db	1021	AGCCCTTCCGCAAGCAGAACCCCGACATCTGATCTACAGTACATGAGCAGCCTGTAGC	1080
Best Local Similarity	92.9%;	Pred. No. 1e-199;		Qy	1232	TGGGCGAGCGACTGTGAGATCGGCGCAGCACCGCGCCCAAGATCGAGGAGCTGGCGAAGCACC	1291
Matches 2147;	Conservative	0;	Mismatches 141;	Db	1081	TGGGCGAGCGACTGTGAGATCGGCGCNGCACCGCACCAAGATCGAGGAGCTGGCGCAGCACC	1140
Qy	170	GCGGCAAGGAGGCCACCCAGATGAAGGACTGCACCGAGCGCCAGCGCAACTTCTTCGCGC	229	Qy	1292	TGCTGCGTGGGGTTTACCAACCCCGCAGACCAAGAGCACCAAGAGAGCGCCCTTCTCTGC	1351
Db	1	GCGGCGGGAAGGACCAACAAATGAAGATTTGACCTGAGAGAGAGCTTAATTTCTTCGCGC	60	Db	1141	TGCTGCGTGGGGTTTACCAACCCCGCAGAACAGACACAGAGAGAGCGCCCTTCTCTGT	1200
Qy	230	AGGAGCTGGCTTCCCGCAGGCGCAAGCGCGAGTTCGCCAGCGAGCAGAACCGCGCCA	289	Qy	1352	CCAT-----CGAGCTGCAACCCCGCAGACCGCTGCAGCCCATCGAGCTGCCCGGAGA	1405
Db	61	AGGAGCTGGCTTCTGTCAGGCGCAAGCGCGAGTTCAGCAGCAGCAGCAGCAGCGCGCA	120	Db	1201	GGATGGGCTAGAGCTGCAACCCCGCAAGTGCAGCCCATATGCTGCCCGGAGA	1260
Qy	290	ACAGCCCCACACCGCGAGCTGCAAGTTCGCGGGG-----ACAAACCCCGCAGCGAGG	343	Qy	1406	AGGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCC	1465
Db	121	ACAGCCCCACCGCGAGCTGCAAGTTCGCGGGG-----ACAAACCCCGCAGCGAGG	180				

```

1261 AGGACAGTGGACCGTGAACACATCCAGAGCTGGTGGCAAGCTGAATCGGCCAGCC 1320
1466 AGATCTACCCCGGATCAAGGTGCGCAGCTGTCAAGCTGTCTCGCGGGGCCAAGGCC 1525
1321 AGATCTAGCGCGGATCAAGGTGAAGCTGTCTCAAGCTGTCTCGCGGGGCCAAGGCC 1380
1526 TGACCGACATGTGCTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCAGA 1585
1381 TGACCGAGGTGATCCCTCTGACGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCAGA 1440
1586 TCCTGCGGAGCCGCTGACCGGCTGTACTAGACCCAGCAGGAGCTGTGTGCGCCAGA 1645
1441 TCCTGAAGAGCCGCTGACGAGGTGTACTAGCAGCCCGCAGCAGGAGCTGTGTGCGCAGA 1500
1646 TCAGAGAGCGGGCCAGCAGCAGTGGACCTACAGATCTACAGGAGCCCTTCAAGAAC 1705
1501 TCAGAGAGCGGGCCAGGCGCAGTGGACCTACAGATCTACAGGAGCCCTTCAAGAAC 1560
1706 TGAAGACCGGCAAGTACCGCAAGATGCGCACCGGCCACCAACGACGTGAAGCAGCTGA 1765
1561 TGAAGACCGGCAAGTACCGCAGTGCAGCGCGGCCACCAACGACGTGAAGCAGCTGA 1620
1766 CCGAGGCGGTGCAGAGATCCCATGAGAGGACCTGCTGAGACCTGCTGGGCGAAGCCCCAAGT 1825
1621 CCGAGGCGGTGCAGAGGTGAGCACCGAGAGCATCGTGTATCTGGGGCAAGATCCCCAAGT 1680
1826 TCCTGCTGCCCATCCAGAGGAGACCTGGGAGACCTGCTGAGACCTGCTGGGCGAGGCCA 1885
1681 TCAAGCTGCCCATCCAGAGGAGACCTGGGAGGCTGCTGAGTGTGAGTACTGGCAGGCCA 1740
1886 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCTGTAAGCTGTGTATACAGC 1945
1741 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCTGTAAGCTGTGTATACAGC 1800
1946 TGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTAGTGGAGCGGCGCGGCCAACCGCG 2005
1801 TGGAGAGGAGCCCATCGTGGCGCGGAGACCTTCTAGTGGAGCGGCGCGGCCAACCGCG 1860
2006 AGACCAAGATCGGCAAGCGGCTACGTGACCGACCGGCGCGGCCGAGAGATCGTGAAGC 2065
1861 AGACCAAGTGGGCGGAGCGGCTACGTGACCGACCGGCGCGGCCGAGAGTGTGTAGCA 1920
2066 TGACCGAGACCAACCAAGAGACCGAGTGCAGGCGCATCCAGCTGGCCCTGCGAGGACA 2125
1921 TCGCCGACACCAACCAAGAGACCGAGTGCAGGCGCATCCAGCTGGCCCTGCGAGGACA 1980
2126 CGGCGAGGAGTGAAACATCGTGAACCGACCGAGTACGCGCTGGGCGATCATCCAGGCC 2185
1981 CGGCGCTGGAGGTGAACATCGTGAACCGACCGAGTACGCGCTGGGCGATCATCCAGGCC 2040
2186 AGCCCGACAGAGCGAGCGGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGG 2245
2041 AGCCCGACAGAGCGAGCGGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGG 2100
2246 AGAAGGTGTACCTGAGCTGGGTGCGCCCGCCCAAGGCGCATCGGCGCAACGAGCAGATCG 2305
2101 AGAAGGTGTACCTGAGCTGGGTGCGCCCGCCCAAGGCGCATCGGCGCAACGAGCAGTGG 2160
2306 ACAAGCTGTGAGCAAGGCGATCCGCAAGGTGTCTTCTGAGCGGCGATCGATGGCGGCA 2365
2161 ACAAGCTGTGAGCGCGCGGATCCGCAAGGTGTCTTCTGAGCGGCGATCGATGGCGGCA 2220
2366 TCCTGATCTACAGTACATGAGCAGCTGTACGTGGCGGCGGCGGCTAGGATCGATT 2425
2221 TCCTGATCTACAGTACATGAGCAGCTGTACGTGGCGGCGGCGGCTAGGATCGATT 2280
2426 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457
2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

```

RESULT 13

```

AR373389
LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2312)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like
particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
FEATURES
Location/Qualifiers
1..2312
source /organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
Best Local Similarity 92.9%; Pred. No. 1e-199;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

Qy 170 GCGGCAAGGAGGCGCACCATGAAAGACTGCAACCGAGCCGAGCCAGGCAACTTCTTCGCG 229
Db 1 GCGGCGCGGAGGAGCACCAATGAAAGATTGCACTGAGAGACAGGCTAAATTTCTTCGCG 60

Qy 230 AGNACCTGGCTTCCCCAGGCGCAAGCCCGCGAGTTCCCCAGGAGCAGACCGCGCCA 289
Db 61 AGNACCTGGCTTCTCTGAGGCGCAAGCCCGCGAGTTCCAGCAGCAGACCGCGCGCCA 120

Qy 290 ACAGCCCCACAGCGCGAGCTGCAGGTGCGCGCGG-----ACAAACCCCGCAGCGAGG 343
Db 121 ACAGCCCCACCGCGCGAGCTGCAGGTGTGGGCGCGGAGACACAGCTGAGCGAGG 180

Qy 344 CCGGCGCGAGCGCCAGGCGACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGC 397
Db 181 CCGGCGCGAGCGCCAGGCGACCGTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC 240

Qy 398 GCGGCTGTGTAGCATCAAGTGGGCGCGCAGATCAAGGAGGCGCTCTGTGACACCGCGCG 457
Db 241 GCGGCTGTGTAGCATCAGGATCGGCGCGCAGCTCAAGGAGGCGCTCTGTGACACCGCGCG 300

Qy 458 CCGACGACACCGTGTGTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTGTGTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCG 360

Qy 518 GCGGCTGTGTAGCATCAAGTGGCGCGCAGTACGACCGAGATCTCTGATCGAGATCTGCG 577
Db 361 GCGGCTGTGTAGCATCAAGTGGCGCGCAGTACGACCGAGATCTCTGATCGAGATCTGCG 420

Qy 578 GCAAGAGGCGCATCGGCGCGCTGTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCA 637
Db 421 GCGCAAGGCGCATCGGCGCGCTGTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCA 480

Qy 638 ACATGCTGACCGAGCTGGGCTGCACCTGAACTTCCCCCATCAGCCCATCGAGACCGTGC 697
Db 481 ACCTGTGACCGAGATCGGCTGCACCTGAACTTCCCCCATCAGCCCATCGAGACCGTGC 540

Qy 698 CCGTGAAGCTGAAGCGCGGCGATGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
Db 541 CCGTGAAGCTGAAGCGCGGCGATGGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600

Qy 758 AGAAGATCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817
Db 601 AGAAGATCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660

Qy 818 AGATCGGCGCGCGAGAACCCCTTACAACACCCCGGTGTTGCGCATCAAGAGAGGAGCA 877
Db 661 AGATCGGCGCGCGAGAACCCCTTACAACACCCCGGTGTTGCGCATCAAGAGAGGAGCA 720

Qy 878 CCAAGTGGCGCAAGCTGGTGGAGCTTCCGCGGAGCTGAACCAAGCGCGCACCGAGCTTCTGGG 937

```

Db 721 CCAAGTGGGCAAGCTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAGAGCTTCTGGG 780  
QY 938 AGGTGAGTGGGCATTCGCCACCCCGCGGCTTGAAGAAGAGAGCTGACCGTGC 997  
Db 781 AGGTGAGTGGGCATTCGCCACCCCGCGGCTTGAAGAAGAGAGCTGACCGTGC 840  
QY 998 TGGAGCTGGGCGACCGCTACTTTCAGGTGCGCCCTGGAGGAGCTTCCGCAAGTACACGG 1057  
Db 841 TGGAGCTGGGCGACCGCTACTTTCAGGTGCGCCCTGGAGGAGCTTCCGCAAGTACACCG 900  
QY 1058 CTTTCACCATCCCAAGCATCAACAACGAGACCCCGCGCATCCGCTACCAAGTACAACTGTC 1117  
Db 901 CTTTCACCATCCCAAGCATCAACAACGAGACCCCGCGCATCCGCTACCAAGTACAACTGTC 960  
QY 1118 TGCCCCAGGGCTGGAAGGCGACGCCAGCATCTTCCAGAGCAGCATGACCAAGATTCCTGG 1177  
Db 961 TGCCCCAGGGCTGGAAGGCGACGCCCGGCATCTTCCAGAGCAGCATGACCAAGATTCCTGG 1020  
QY 1178 AGCCCTTCGGCGCCGCAACCCGAGATCGTGATCTACCA-----GGCCCCCTGTACG 1231  
Db 1021 AGCCCTTCGGCAAGCAGAACCCCGCATCGTGATCTACCAAGTACATGAGACCTGTACG 1080  
QY 1232 TGGGAGGAGCACTGAGATCGGCCAGCACCGCGCAAGATCGAGAGCTGGCGAAGCACCC 1291  
Db 1081 TGGGAGGAGCACTGAGATCGGCCAGCACCGCGCAAGATCGAGAGCTGGCGAAGCACCC 1140  
QY 1292 TGCTCGCTGGGGCTTCCACACCCCGCAAGAGCACCAAGAGAGCCCGCTTCCTGTC 1351  
Db 1141 TGCTCGCTGGGGCTTCCACACCCCGCAAGAGCACCAAGAGAGCCCGCTTCCTGTC 1200  
QY 1352 COAT-----CGAGCTGCACCCGCAAGTGGACCGTGCAGCCCATCGAGTGCCTCGAGA 1405  
Db 1201 GGATGGGCTACGAGTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGTGCCTCGAGA 1260  
QY 1406 AGGAGAGTGGACCGCTGAACACATCCAGAACTGCTGGGCAAGCTGAACCTGGGCGAGCC 1465  
Db 1261 AGGAGAGTGGACCGCTGAACACATCCAGAACTGCTGGGCAAGCTGAACCTGGGCGAGCC 1320  
QY 1466 AGATCTACCGGCGATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGGCGCAAGGCGCC 1525  
Db 1321 AGATCTACCGGCGATCAAGGTGGAAGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCGCC 1380  
QY 1526 TGACCGACATCGTGCCTGACCGAGGAGGCGGAGCTGAGCTGCGCGAGAACCGCGAGA 1585  
Db 1381 TGACCGAGGTGATTCCTCCCTGACCGAGGAGGCGGAGCTGAGCTGCGCGAGAACCGCGAGA 1440  
QY 1586 TCCTCGGAGCGCCGTGCAACGCGGTACTACGACCCCGCAAGGACCTGCTGGGCGGAG 1645  
Db 1441 TCCTGAAGGAGCCGTGCACGAGGTGTACTAGACCCCGCAAGGACCTGCTGGGCGGAG 1500  
QY 1646 TCCAGAGCGGGCCACGACCAAGTGGACTACAGATCTACAGAGGCGCTTCAAGAACCC 1705  
Db 1501 TCCAGAGCGGGCCACGACCAAGTGGACTACAGATCTACAGAGGCGCTTCAAGAACCC 1560  
QY 1706 TGAACGCGGCAAGTACCCCAAGATGCCACCGCGCCACCAACACAGCTGAGCAGCTGA 1765  
Db 1561 TGAACGCGGCAAGTACCCCGCATGCGCGGCGCCACCAACACAGCTGAGCAGCTGA 1620  
QY 1766 CCGAGGCGGTGAGAAAGATCGCCATGAGAGCATCGTGATCTGGGGCGAAGACCCCAAGT 1825  
Db 1621 CCGAGGCGGTGAGAAAGTGAACCCGAGAGCATCGTGATCTGGGGCGAAGATCCCAAGT 1680  
QY 1826 TCCGCTGCGCCATCCAGAAAGAGACTGGGAGACTGTGGACCGACTTGGGAGGCGCA 1885  
Db 1681 TCAAGCTGCGCATCCAGAAAGAGACTGGGAGGCGCTGGTGGATGAGTACTGGCAGGCGCA 1740  
QY 1886 CTTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCTGAGAGCTGTGTTACAGC 1945  
Db 1741 CTTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCTGAGAGCTGTGTTACAGC 1800  
QY 1946 TGGAGAGAGGCGCCATCATCGGCGCGGAGACCTTCTAGTGTGACGCGCGCCCAACCGCG 2005

Db 1801 TGGAGAGGAGCCCATCGTGGGCGCGAGAGCTTCTACGTGGACGCGCGCCCAACCGCG 1860  
QY 2006 AGACCAAGATCGGCAAGCCCGGCTACGTGACCGGCGCGGCGCGGAGAGATCGTGAGCC 2065  
Db 1861 AGACCAAGTCTGGCAAGCCCGGCTACGTGACCGGCGCGGCGCGGAGAGTGTGTAGCA 1920  
QY 2066 TGACCGAGACCAACCAACAGAAAGCCGAGCTGCAAGGCGCATCCAGCTGGGCCCTGCAGGACA 2125  
Db 1921 TCGCGGACACCAACCAAGAAAGCCGAGCTGCAAGGCGCATCCACTGGGCCCTGCAGGACA 1980  
QY 2126 CGCGAGCGAGGTGAACATCGTGACCGGACAGCCAGTACGCCCTGGGCATCATCCAGGCC 2185  
Db 1981 CGGCGCTGGAGGTGAACATCGTGACCGGACAGCCAGTACGCCCTGGGCATCATCCAGGCC 2040  
QY 2186 AGCCCGAAGAGCGAGAGCGAGCTGCTGTAACCAAGATCATCGAGAGCTGATCAAGAGG 2245  
Db 2041 AGCCCGAAGAGCGAGAGCGAGCTGCTGTAACCAAGATCATCGAGAGCTGATCAAGAGG 2100  
QY 2246 AGAAGGTGTACTCGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCAACGAGCAGATCG 2305  
Db 2101 AGAAGGTGTACTCGGCTGGGTGCCCGCCCAAGGGCATCGCGGCAACGAGCAGGTGG 2160  
QY 2306 ACAAGCTGGTGAGCAAGGCGCATCCGCAAGGTGCTGTTCTCGAGCGGCATCGATGCGGCA 2365  
Db 2161 ACAAGCTGGTGAGCGCGGCATCCGCAAGGTGCTGTTCTCGAACGGCATCGATGCGGCA 2220  
QY 2366 TCGTGATCTACAGTACATGGAACGACCTGTACCTGGGAGCGGCGGCGCTAGGATCGATT 2425  
Db 2221 TCGTGATCTACAGTACATGGAACGACCTGTACCTGGGAGCGGCGGCGCTAGGATCGATT 2280  
QY 2426 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 14  
AX427936  
LOCUS AX427936 9788 bp DNA linear PAT 20-JUN-2002  
DEFINITION Sequence 174 from Patent WO0232943.  
ACCESSION AX427936  
VERSION AX427936.1 GI:21538023  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE 1  
AUTHORS Huang,Y. and Nabel,G.J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 174 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES  
source Location/Qualifiers  
1. .9788  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Plasmid pVR1012x/s containing HIV genes"

ORIGIN  
Query Match 81.1%; Score 1992.4; DB 6; Length 9788;  
Best Local Similarity 91.2%; Pred. No. 2.9e-199;  
Matches 2181; Conservative 0; Mismatches 181; Indels 30; Gaps 5;  
QY 14 TGGCGGAGGCCATGAGCCAGGCCACCGCGCAACATCCTGATGAGCGGCGAGCACTTCA 73  
Db 2961 TGGCGGAGGCCATGAGCAAGGTGAACACCAACATCATGATGAGCGGCGAGCACTGCA 3020  
QY 74 AGGCCCCCAAGCGGATCATCAAGTGTCTCAACTGCGCAAGAGGGCCACATCGGCCGCA 133  
Db 3021 AGGCCCCCAAGCGGACCATCAAGTGTCTCAACTGCGGCAAGAGGGCCACCTGGGCCGCA 3080  
QY 134 ACTGCGCGCCCCCGCGCAAGAGGGCTGCTGGAAGTGTGCGCAAGAGGGCCACCAAGATGA 193



Db 3081 ACTCGCGGCCCCCGCAAGAGGGCTGCTGGAAAGTGGCGCAAGAGGGCCACAGGTGA 3140  
Qy 194 AGGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCCCAGGGCA 253  
Db 3141 AGGACTGACCGAGCGCAAGGCTAA-----TAGGGAAGATCTGGCCTTCCCAAGGGA 3194  
Qy 254 AGGCCGGGAGTTTCCAGAGGAGAGAAACCGGGCAACAGGCCCAACAGCGCGAGTGC 313  
Db 3195 AGGCCAGGGAAATTTCTTCAAGAGCAGACCAAGAGCAACAGCCCCCAAGAGAGAGCTTC 3254  
Qy 314 AGGTGCGCGG-----CGACAACCCCGCAGCGAGGCGCGCGCGAGCGCCAGGGCA--- 364  
Db 3255 AGGTTTGGGAGAGACAACAACTTCCCTCTCAGAAAGCAGAGGCCGATAGACAAGAACTG 3314  
Qy 365 ---CCCTGAACTTCCCCCAGATCACTCTGTGGCAGCGCCCCCTGGTGAGCATCAAGTGG 421  
Db 3315 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAATAAAGATAG 3374  
Qy 422 GCGGCCAGATCAGGAGGCCCTGTGGACACCGCGCGCGAGACAACCGTGTGGAGAGA 481  
Db 3375 GGGGCCAGCTGAAGAGGCCCCCTTAGACACCGCGCGCGAGACAACCGTGTGGAGAGA 3434  
Qy 482 TGAGCCTGCCCGCAAGTGGAAAGCCCAAGATGATCGGGCGCATCGGGCGCTTCAATCAAG 541  
Db 3435 TGAACCTGCCCGCGCTGGAAGCCCAAGATGATCGGGCGCATCGGGCGCTTCAATCAAG 3494  
Qy 542 TGGCCCAAGTACGACAGATCTCTGATCGAGATCTCGGCAAGAGGCCATCGGCAACCGTGC 601  
Db 3495 TGGGCCAGTACGACAGATCTCTGATCGAGATCTCGGGCCACAGGCCATCGGCAACCGTGC 3554  
Qy 602 TGATCGGCCCAACCGCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGTGCA 661  
Db 3555 TGGTGGGCCCCACCCCGTGAACATCATCGGCGCGCAACCTGCTGACCCAGATCGGCTGCA 3614  
Qy 662 CCCTGAATTTCCCATCAGCCCCATCGAGCCGTCAGACCGTGCAGCTGAAGCCCGGATGG 721  
Db 3615 CCCTGAATTTCCCATCAGCCCCATCGAGCCGTCAGACCGTGCAGCTGAAGCCCGGATGG 3674  
Qy 722 ACGGCCCCAAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGACGCCCA 781  
Db 3675 ACGGCCCCAAGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTGAGTGA 3734  
Qy 782 TCTGCGAGGAGATGGAGAAGAGGGCAAGATTCACAAAGATCGGCCCCCGAGAACCCCTTACA 841  
Db 3735 TCTGCAACGAGATGGAGAAGAGGGCAAGATTCAGCAAGATCGGCCCCCGAGAACCCCTTACA 3794  
Qy 842 ACAACCCCGTTCGCGCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGGTGAAT 901  
Db 3795 ACAACCCCGTTCGCGCATCAAGAAAGAGAGACAGACCAAGTGGCGCAAGCTGGTGAAT 3854  
Qy 902 TCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGGAGGTGACAGTGGGCATCCCCCACC 961  
Db 3855 TCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGGAGGTGACAGTGGGCATCCCCCACC 3914  
Qy 962 CCGCGGCTCTGAAGAAGAGAGCGTGACCGTGTGGACGTGGGCGACGCCCTACTTCA 1021  
Db 3915 CCGCGGCTCTGAAGAAGAGAGCGTGACCGTGTGGACGTGGGCGACGCCCTACTTCA 3974  
Qy 1022 CGGTGCCCTTGGACGAGGACTTCGGAAGTACACCGCTTCAACATCCCGAGATCAACA 1081  
Db 3975 GCGTGCCCTTGGACAAGGACTTCGGAAGTACACCGCTTCAACATCCCGAGATCAACA 4034  
Qy 1082 ACAGAGCCCCCGGCATCCGTTACAGTACACGTTGCTGCCCGCAGGCTGGAAGGGCAGCC 1141  
Db 4035 ACAGAGCCCCCGGCATCCGTTACAGTACACGTTGCTGCCCGCAGGCTGGAAGGGCAGCC 4094  
Qy 1142 CCAGCATTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201  
Db 4095 CCGCCATTTCCAGTCAGCATGACCAAGATCTCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154  
Qy 1202 AGATCGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCATCGGATCGGCC 1255  
Db 4155 ACATCGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCATCGGATCGGCC 4214

Qy 1256 AGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTCCACCAACC 1315  
Db 4215 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCGTGGGGCTTCCACCAACC 4274  
Qy 1316 CCGACAAGAGCACCAAGAGAGGCCCTTCTTCTGCCCCAT-----CGAGTGCACCCCG 1369  
Db 4275 CCGACAAGAGCACCAAGAGAGGCCCTTCTTCTGTTGGATGGGCTACGAGTGCACCCCG 4334  
Qy 1370 ACAAGTGCACCGTGCAGCCCATCGAGCTGCCGAGAAGGAGAGCTGGACCGTGAACGACA 1429  
Db 4335 ACAAGTGCACCGTGCAGCCCATCGTCTGCCGAGAAGGAGAGCTGGACCGTGAACGACA 4394  
Qy 1430 TCAGAAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCCGCATCAAGGTGC 1489  
Db 4395 TCAGAAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTTACGCCCGCATCAAGGTGC 4454  
Qy 1490 GCAGCTGTGCAAGCTGTCTGCGCGCGCCAAAGCCCTTGAACGACATCGTGCCTTGACCG 1549  
Db 4455 GCCAGCTGTGCAAGCTGTCTGCGCGGCAACCAAGGCCCTTGAACGAGGTGTGCCCCTGACCG 4514  
Qy 1550 AGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGCG 1609  
Db 4515 AGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGAAAGAGCCCGTGCACGCG 4574  
Qy 1610 TGTACTACGACCCCGCAGCAAGGACCTGTGTGGCCGAGATCCAGAAAGCAGGGCCACACCAAGT 1669  
Db 4575 TGTACTACGACCCCGCAGCAAGGACCTGTATCGCCGAGATCCAGAAAGCAGGGCCACGCGCAGT 4634  
Qy 1670 GGAACCTACAGATCTACAGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCAAGA 1729  
Db 4635 GGAACCTACAGATCTACAGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCGCA 4694  
Qy 1730 TGGCGCACCGCCCAACACGAGCTGAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCA 1789  
Db 4695 TGAAGGGCGCCCAACACGAGCTGAAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCA 4754  
Qy 1790 TGGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTTCCGCTTCCCATCCAGAAAGAGA 1849  
Db 4755 CCGAGAGCATCTGTATCTGGGCAAGACCCCAAGTTTCAAGTTCAGCTGCCCATCCAGAAAGAGA 4814  
Qy 1850 CTTGGGAGACCTGGTGAACCGACTATCTGGCAGGCGACCTGGATCCCGAGTGGAGTTGC 1909  
Db 4815 CTTGGGAGGCTGGTGGACCGAGTACTTGGCAGGCGCACTGGATCCCGAGTGGAGTTGC 4874  
Qy 1910 TGAACACCCCGCTGGTGAAGCTGTGTGTAACAGCTGGAGAGGAGCCCATCATCGCG 1969  
Db 4875 TGAACACCCCGCTGGTGAAGCTGTGTGTAACAGCTGGAGAGGAGCCCATCATCGCG 4934  
Qy 1970 CCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGCT 2029  
Db 4935 CCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGTGGGCGAGGCCCGCT 4994  
Qy 2030 ACCTGACCGACCGGGCGCGCAGAAAGATCGTGAACCTGACCGAGACCAACCAACAGAGA 2089  
Db 4995 ACCTGACCGACCGCGCGCGCAGAAAGTGTGCTGCTGACCGCACCAACCAACAGAGA 5054  
Qy 2090 CCGAGCTGACGGCCATCCAGCTGGCCCTGCAGACAGCGGCGAGGAGTGAACATCTGTA 2149  
Db 5055 CCGAGCTGACGGCCATCCACCTGGCCCTGCAGACAGCGGCCCTGGAGGTGAACATCTGTA 5114  
Qy 2150 CCGACAGCCAGTAGTCCGCTGGGCATCATCCAGCGCCAGCCGCAAGAGCGAGAGCGAGC 2209  
Db 5115 CCGACAGCCAGTAGTCCGCTGGGCATCATCCAGCGCCAGCCGCAAGAGCGAGAGCGAGC 5174  
Qy 2210 TGGTGAACAGATCATCGAGCAGCTGATCAAGAAAGAGAAAGGTGTACCTGAGCTGGGTGC 2269  
Db 5175 TGGTGAACAGATCATCGAGCAGCTGATCAAGAAAGAGAAAGGTGTACCTGAGCTGGGTGC 5234  
Qy 2270 CCGCCCAAGAGGCGATCGGCGCAACGAGCAGATTCGACAGCTGGTGGAGCAAGGGCATCC 2329  
Db 5235 CCGCCCAAGAGGCGATCGGCGCAACGAGCAGGTGGAACCGCCCTGGTGGAGCGCGGCATCC 5294



Qy	2330	GCAAGTGTCTTTCTGGACCGCATCGATGCGGCATCGTGATCTACCAGTA	2381
Db	5295	GCAAGGTGTCTTTCTGGACCGCATCGAACAAGCCGAGGACGAGAA	5346
RESULT 15			
LOCUS	AX427931	9169 bp DNA linear	PAT 20-JUN-2002
DEFINITION	Sequence 169 from Patent W00232943.		
ACCESSION	AX427931		
VERSION	AX427931.1	GI:21538018	
KEYWORDS	. synthetic construct synthetic construct other sequences; artificial sequences.		
SOURCE	Huang, Y. and Nabel, G.J. Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization Patent: WO 0232943-A 169 25-APR-2002; GOVERNMENT OF THE UNITED STATES (US)		
ORGANISM	Location/Qualifiers 1. .9169 /organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="plasmid pVR1012x/s containing HIV genes"		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match	80.5%; Score 1978.4; DB 6; Length 9169;		
Best Local Similarity	91.1%; Pred. No. 8.5e-198;		
Matches 2178; Conservative	0; Mismatches 186; Indels 28; Gaps 6;		
Qy	14	TGSCCGAGGCCATGACCGCAGGCCACCGGCCAACATCCTGTATGCGCGCAGCAACTTCA	73
Db	2967	TGSCCGAGGCCATGACCGCAGGCCATGACCAACATCATGATGAGCGCGGCAACTTCA	3026
Qy	74	AGSGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGGCCACATCGCCGCA	133
Db	3027	AGSGCCAGAAGCCAT---CAAGTGTCTCACTGCGGCAGGAGGGCCACTCTGCCCGCA	3083
Qy	134	ACTGCGCGCCCCGCCAAGAAAGGGTGTGTGAAGTGGCGCAAGAGGGGCCACACAGATGA	193
Db	3084	ACTGCGCGCCCCGTGCAAGAGGGTGTGTGAAGTGGCGCAAGAGGGGCCACACAGATGA	3143
Qy	194	AGGACTGCACCGACCGCAGGCCNACTTCTTCGCGAGGACCTGSCCTTCCCAGGGCA	253
Db	3144	AGGACTGCACCGACCGACAGGGCTAA-TTTTTTAGGGAAGATCTGGCCCTTCCCAAGGGA	3202
Qy	254	AGGCGCGAGTTTCCCGACGAGCAGAACCGCGCCAAGCCCCACACGCGCGAGCTGC	313
Db	3203	AGGCGAGGAATTTTCTTACAGAGCAGACCAAGCCCAACGCCCCACCAAGAGAGAGCTTC	3262
Qy	314	AGGTGCGCGG-----CGACAACCCCCCGACGAGGCCCGCGCGCGAGCGCCAGGGCA---	364
Db	3263	AGGTTTGGGGAAGAGACAACATCCCTCTCAGAAGCAGGAGCGCATAGACAAGAACTG	3322
Qy	365	----CCCTGAACATTCCCCHAGATCACCTGTGCGACGGCCCCCTGTGTAGCATCAAGGTGG	421
Db	3323	TATCTCTTAGCTTTCCCTCTAGATCACTCTTTTGGCAGCGACCCCTCGTCAATAAAGATAG	3382
Qy	422	CGSGCCAGATCAAGGAGGCCCTGTGACACCGCGCGCGCAGACACCGTGTGGAGGAGA	481
Db	3383	GGGGCCAGCTGAAGGAGGCCCTTTAGACACCGCGGCCGACACACCTGTGTGGAGAGA	3442
Qy	482	TGAGCCTGCCCGCAAGTGAAGCCCAAGATGATCGGGGCATCGCGGCTTCTATCAAGG	541
Db	3443	TGAACCTGCCGCGCGCTGGAAGCCCCAAGATGATCGGGGCATCGGGGCTTCTATCAAGG	3502
Qy	542	TGGCCNGTACGACCAAGATCTTGATCGAGATCTGGCGGCAGAGGCCATCGGCAACCGTGC	601
Db	3503	TGGGCCAGTACGACCGATCTTGATCGAGATCTGGGGCCACAAGGCCATCGGCAACCGTGC	3562

```
Db 4643 | 4643 | GGACCTACAGATCTACAGAGCCCTCAAGAACCTGAAGACCGGAAGTACGCCCGCA 4702
Qy 1730 | 1730 | TGGCAGCCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCCGTGCAGAAAGATCGCCA 1789
Db 4703 | 4703 | TGAAGGGCGGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCAGAGATCGCCA 4762
Qy 1790 | 1790 | TGGAGAGATCGTGATCTGGGGCAAGACCCCAAGTTCCGGCTGCCCATCCAGAAGGAGA 1849
Db 4763 | 4763 | CCGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCATCCAGAAGGAGA 4822
Qy 1850 | 1850 | CCTGGGAGACTGCTGACCGACTACTGGCAGAGCCACTGGATCCCGAGTGGGAGTTGG 1909
Db 4823 | 4823 | CTTGGGAGGCTGCTGACCGAGTACTGGCAGGCCACTGGATCCCGAGTGGGAGTTGG 4882
Qy 1910 | 1910 | TGAACACCCCCCTCGTGAAGCTGTGGTACCAAGCTGGGAAGAGGCCCATCATCGGCG 1969
Db 4883 | 4883 | TGAACACCCCCCTCGTGAAGCTGTGGTACCAAGCTGGGAAGAGGCCCATCATCGGCG 4942
Qy 1970 | 1970 | CCGAGACTTCTACGTGGAAGGCCCGCCCAACCGCGAGACCAAGATCGGCAAGGCCGGCT 2029
Db 4943 | 4943 | CCGAGACTTCTACGTGGAAGGCCCGCCCAACCGCGAGACCAAGTGGGCAAGGCCGGCT 5002
Qy 2030 | 2030 | ACGTGACCGAGCCGGCGCGCAGAGATCGTGAGCCCTGACCGAGACCAACCAACGAGA 2089
Db 5003 | 5003 | ACGTGACCGAGCCGGCGCGCAGAGATCGTGAGCCCTGACCGAGACCAACCAACGAGA 5062
Qy 2090 | 2090 | CCGAGCTGCAGGCCCATCCAGCTGGCCCTGCAGGACAGCGGAGGAGTGAACATCGTGA 2149
Db 5063 | 5063 | CCGAGCTGCAGGCCCATCCAGCTGGCCCTGCAGGACAGCGGCTGGAGGTGAACATCGTGA 5122
Qy 2150 | 2150 | CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAAGCCGACAGAGCGAGAGCGAGC 2209
Db 5123 | 5123 | CCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAAGCCGACAGAGCGAGAGCGAGC 5182
Qy 2210 | 2210 | TGGTGAACCCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGATCCTGAGCTGGGTGC 2269
Db 5183 | 5183 | TGGTGAACCCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGATCCTGAGCTGGGTGC 5242
Qy 2270 | 2270 | CCGCCCAAGGGGATCGGGGCAACGAGCAGATCGACCAAGCTGGTGAAGGAGGATCC 2329
Db 5243 | 5243 | CCGCCCAAGGGGATCGGGGCAACGAGCAGGTGGACGGCTGGTGAAGGAGGATCC 5302
Qy 2330 | 2330 | GCAAGGTGCTTCTCGGAGCGGCATCGATGGCGGCATCGTGATCTACAGTA 2381
Db 5303 | 5303 | GCAAGGTGCTTCTCGGAGCGGCATCGATGGCGGCATCGATGGCGGCATCGAGAA 5354
```

Search completed: June 1, 2005, 20:20:49  
Job time : 10525 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1264.25 Seconds  
(without alignments)  
11504.726 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457

Sequence: 1 gtcagccaccatggccga.....gggctagaccggatgaattc 2457

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

- 1: Geneseqn1980s:\*
- 2: Geneseqn1990s:\*
- 3: Geneseqn2000s:\*
- 4: Geneseqn2001as:\*
- 5: Geneseqn2001bs:\*
- 6: Geneseqn2002as:\*
- 7: Geneseqn2002bs:\*
- 8: Geneseqn2003as:\*
- 9: Geneseqn2003bs:\*
- 10: Geneseqn2003cs:\*
- 11: Geneseqn2003ds:\*
- 12: Geneseqn2004as:\*
- 13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	100.0	2457	6	ABL39961
2	2457	100.0	2457	12	ADM73766
3	2443.4	99.4	2445	8	ACA03546
4	2443.4	99.4	2445	10	ADCI3264
5	2436.4	99.2	3930	10	ADCI3230
6	2436.2	99.2	2463	6	ABL39960
7	2436.2	99.2	2463	12	ADM73765
8	2434.8	99.1	3930	10	ADCI3231
9	2434.8	99.1	3930	10	ADCI3232
10	2434.8	99.1	5184	8	ACA03591
11	2434.8	99.1	5184	10	ADCI3279
12	2428.6	98.8	2457	8	ACA03547
13	2428.6	98.8	2457	10	ADCI3265
14	2415.4	98.3	2469	6	ABL39959
15	2415.4	98.3	2469	12	ADM73764
16	2404.4	97.9	3531	8	ADCI3234
17	2403.4	97.8	2457	8	ACA03548
18	2403.4	97.8	2457	10	ADCI3266
19	2402.8	97.8	3537	10	ADCI3236
20	2401.8	97.8	5145	8	ACA03521

21	2401.8	97.8	5145	10	ADCI3233	Adcl3233 DNA of HI
22	2391.8	97.3	3538	10	ADCI3235	Adcl3235 DNA of HI
23	2343.4	95.4	3607	8	ACA03551	Aca03551 Synthetic
24	2343.4	95.4	3607	10	ADCI3269	Adcl3269 DNA of HI
25	2325.2	94.6	3597	8	ACA03549	Aca03549 Synthetic
26	2325.2	94.6	3597	10	ADCI3267	Adcl3267 DNA of HI
27	2295.8	93.4	3624	8	ACA03550	Aca03550 Synthetic
28	2295.8	93.4	3624	10	ADCI3268	Adcl3268 DNA of HI
29	2136	86.9	2460	8	ACA03541	Aca03541 Synthetic
30	2136	86.9	2460	8	ACC78505	Acc78505 HIV p2Pol
31	2135	86.9	3564	8	ACC78488	Acc78488 HIV GagPo
32	2135	86.9	3564	8	ACC78489	Acc78489 HIV GagPo
33	2134.4	86.9	4716	10	ADCI3238	Adcl3238 DNA of HI
34	2134.4	86.9	4716	10	ADCI3238	Adcl3238 DNA of HI
35	2130.6	86.7	3999	8	ACC78484	Acc78484 HIV GagCo
36	2129	86.7	3999	8	ACC78485	Acc78485 HIV GagCo
37	2129	86.7	3999	8	ACC78486	Acc78486 HIV GagCo
38	2129	86.7	5283	8	ACA03584	Aca03584 Synthetic
39	2129	86.7	5283	8	ACC78529	Acc78529 HIV TatRe
40	2128.8	86.6	4713	8	ACA03592	Aca03592 Synthetic
41	2128.8	86.6	4713	10	ADCI3280	Adcl3280 DNA of HI
42	2128.4	86.6	3462	10	ADCI3237	Adcl3237 DNA of HI
43	2115.2	86.1	2466	8	ACA03542	Aca03542 Synthetic
44	2115.2	86.1	2466	8	ACC78506	Acc78506 HIV p2Pol
45	2103	85.6	3735	8	ACA03545	Aca03545 Synthetic

## ALIGNMENTS

## RESULT 1

ABL39961  
ID ABL39961 standard; DNA; 2457 BP.

XX ABL39961;

XX 15-MAY-2002 (first entry)

XX Synthetic construct PR975YMM SEQ ID NO:32.

XX Human immunodeficiency virus type C; antigenic HIV type C protein; immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; nef; immunostimulant; gene therapy; gene; ds.

XX Human immunodeficiency virus; type C.

OS Synthetic.

XX WO200204493-A2.

XX 17-JAN-2002.

XX 05-JUL-2001; 2001WO-US021241.

XX 05-JUL-2000; 2000US-00610313.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2002-154920/20.

XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.

XX Claim 1; Fig 10; 233pp; English.

XX The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (I). (I) have immunostimulant activity and can be used in gene therapy.

CC The HIV type C polynucleotides are useful in applications including DNA  
CC immunisation, generation of packaging cell lines, and production of HIV  
CC Type C proteins. The polynucleotides are particularly useful in gene  
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and  
CC ABB06204 to ABB06215 represent sequences used in the exemplification of  
CC the present invention  
XX  
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 100.0%; Score 2457; DB 6; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 9.7e-296;  
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAGCCACCATGGCGAGCCATGAGCCAGCCAGCCGCAACATCTGTATGCAG 60  
DB 1 GTCGAGCCACCATGGCGAGCCATGAGCCAGCCAGCCGCAACATCTGTATGCAG 60

QY 61 GCGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGGGCAAGAGGGC 120  
DB 61 GCGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGGGCAAGAGGGC 120

QY 121 CACATGCGCGCAACTGCGCGCCCCCGCGCAAGAAAGGGCTGTGAAAGTGGCGCAAGGAG 180  
DB 121 CACATGCGCGCAACTGCGCGCCCCCGCGCAAGAAAGGGCTGTGAAAGTGGCGCAAGGAG 180

QY 181 GGCCACCAAGATCAAGGATGTCACAGCGCCAGCGCCAACTTCTTCCGCGAGGACCTGGCC 240  
DB 181 GGCCACCAAGATCAAGGATGTCACAGCGCCAGCGCCAACTTCTTCCGCGAGGACCTGGCC 240

QY 241 TTCCCGCCAGGGCAAGGCCCGCGAGTTCCCGAGCAGAGAAACCGCGCCAAACAGCCCGCC 300  
DB 241 TTCCCGCCAGGGCAAGGCCCGCGAGTTCCCGAGCAGAGAAACCGCGCCAAACAGCCCGCC 300

QY 301 AGCCCGAGCTGACAGTGGCGGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCCAG 360  
DB 301 AGCCCGAGCTGACAGTGGCGGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCCAG 360

QY 361 GGCACTCTGAACTTCCCGCAGATCACTGTGGAGCGCCCGCTTCTTGGAGCATCAAGGTG 420  
DB 361 GGCACTCTGAACTTCCCGCAGATCACTGTGGAGCGCCCGCTTCTTGGAGCATCAAGGTG 420

QY 421 GCGGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCAGACACCGTGTCTGAGGAG 480  
DB 421 GCGGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCAGACACCGTGTCTGAGGAG 480

QY 481 ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAG 540  
DB 481 ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAG 540

QY 541 GTGCGCAGTACGACCATCTGTATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
DB 541 GTGCGCAGTACGACCATCTGTATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600

QY 601 CTGATCGGCGCCCGCCCGTGAACATCATCGCGCCCAACATGCTGACCCAGCTGGGCTGC 660  
DB 601 CTGATCGGCGCCCGCCCGTGAACATCATCGCGCCCAACATGCTGACCCAGCTGGGCTGC 660

QY 661 ACCCTGAACTTCCCGCATCAGCGCCCATCGAGACCGTGGCGTGAAGTGAAGCGCGCATG 720  
DB 661 ACCCTGAACTTCCCGCATCAGCGCCCATCGAGACCGTGGCGTGAAGTGAAGCGCGCATG 720

QY 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCGCTGACCGCC 780  
DB 721 GACGCGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCGCTGACCGCC 780

QY 781 ATCTCGAGGAGATGAGAGAGGAGGCGCAAGATCAACAAAGTGGCGCCCGAGAACCCCTAC 840  
DB 781 ATCTCGAGGAGATGAGAGAGGAGGCGCAAGATCAACAAAGTGGCGCCCGAGAACCCCTAC 840

QY 841 AACACCCCGTGTTCGCCATCAAGAAAGAGGACGACCAAGTGGCGCAAGTGTGTGGAC 900  
DB 841 AACACCCCGTGTTCGCCATCAAGAAAGAGGACGACCAAGTGGCGCAAGTGTGTGGAC 900

QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCGCCAC 960  
DB 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCGCCAC 960

QY 961 CCCGCGGCTTGAAGAAGAAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTACTTC 1020  
DB 961 CCCGCGGCTTGAAGAAGAAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTACTTC 1020

QY 1021 AGCGTGCCCTTGGAGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGATCAAC 1080  
DB 1021 AGCGTGCCCTTGGAGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGATCAAC 1080

QY 1081 AACGAGACCCCGCGCATCCGTACAGTACAAAGTGTGCTGCCAGGGTGAAGGGGAGC 1140  
DB 1081 AACGAGACCCCGCGCATCCGTACAGTACAAAGTGTGCTGCCAGGGTGAAGGGGAGC 1140

QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCAACCCC 1200  
DB 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCAACCCC 1200

QY 1201 GAGATCGTGTATCTACAGGCCCCCTGTGTACGTGGGAGGAGCCTGGAGATCGGCCAGC 1260  
DB 1201 GAGATCGTGTATCTACAGGCCCCCTGTGTACGTGGGAGGAGCCTGGAGATCGGCCAGC 1260

QY 1261 CGCGCAAGATCGAGGAGCTGCGAAGCACCTGTGCTGGGCTTCAACACCCCGGAG 1320  
DB 1261 CGCGCAAGATCGAGGAGCTGCGAAGCACCTGTGCTGGGCTTCAACACCCCGGAG 1320

QY 1321 AAGAGCAACCAAGAGGAGCCCCCTTCTGCTCCATCGAGCTGCAACCCCGCAAGTGGAGC 1380  
DB 1321 AAGAGCAACCAAGAGGAGCCCCCTTCTGCTCCATCGAGCTGCAACCCCGCAAGTGGAGC 1380

QY 1381 GTGAGCCCATCGAGCTGCTCCGAGAGAGAGTGGACCTGTGAACGATCCAGAGAGTGTG 1440  
DB 1381 GTGAGCCCATCGAGCTGCTCCGAGAGAGAGTGGACCTGTGAACGATCCAGAGAGTGTG 1440

QY 1441 GTGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGCATCAAGTGTGCGCAGCTGTGC 1500  
DB 1441 GTGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGCATCAAGTGTGCGCAGCTGTGC 1500

QY 1501 AAGCTGTGCGCGCGCCCAAGCGCTGTGACGACATCTGTGCGCCCTGACCGAGAGCGCGAG 1560  
DB 1501 AAGCTGTGCGCGCGCCCAAGCGCTGTGACGACATCTGTGCGCCCTGACCGAGAGCGCGAG 1560

QY 1561 CTGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGTGACGCGGTGTACTACGAC 1620  
DB 1561 CTGAGCTGGCGGAGAACCGCGAGATCTGCGCGAGCCCGTGTGACGCGGTGTACTACGAC 1620

QY 1621 CCCAGAGGAGCTGGTGGCGGAGATCCAGAAAGAGGCGCCACGACAGTGGACCTTACCAG 1680  
DB 1621 CCCAGAGGAGCTGGTGGCGGAGATCCAGAAAGAGGCGCCACGACAGTGGACCTTACCAG 1680

QY 1681 ATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGTGGCGACCGCC 1740  
DB 1681 ATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGTGGCGACCGCC 1740

QY 1741 CACACCAAGCAGTGAAGCAGCTGACCGAGCGCTGTGAAGATCGCCATGGAGAGCATC 1800  
DB 1741 CACACCAAGCAGTGAAGCAGCTGACCGAGCGCGTGTGAAGATCGCCATGGAGAGCATC 1800

QY 1801 GTGATCTGGGCAAGACCCCGCTGCGCTGCGCATCCAGAAAGGAGACCTGGGAGAGC 1860  
DB 1801 GTGATCTGGGCAAGACCCCGCTGCGCTGCGCATCCAGAAAGGAGACCTGGGAGAGC 1860

QY 1861 TGGTGGACCGACTACTGGCAGCGCACCTGTGATCTCCCGAGTGGAGTTCGTGAACACCCCG 1920  
DB 1861 TGGTGGACCGACTACTGGCAGCGCACCTGTGATCTCCCGAGTGGAGTTCGTGAACACCCCG 1920

QY 1921 CCCCTGGTGAAGCTGTGTGTACAGCTGGAGAGGCGCCATCATCGGCGCGAGACCTTC 1980  
DB 1921 CCCCTGGTGAAGCTGTGTGTACAGCTGGAGAGGCGCCATCATCGGCGCGAGACCTTC 1980

QY 1981 TACGTGGAGCGGCGCGCCAAACCGGAGAGACCAAGATCGGCAAGGCGCGGCTACGTGAGCCGAC 2040

Db 1981 TACGTGGACGGCGCCGCAACCGGAGACCAAGATCGCAAGCGCGGTACGTGACCGAC 2040  
 Qy CGGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGACCGAGCTGCGAG 2100  
 Db CGGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGACCGAGCTGCGAG 2100  
 Qy GCATTCACAGCTGGCGCTGACGAGACCGAGCGGCGAGGAGTGAACATCGTGACCGAGCCAG 2160  
 Db GCCATTCACAGCTGGCGCTGACGAGACCGAGCGGCGAGGAGTGAACATCGTGACCGAGCCAG 2160  
 Qy TACGCGCTGGGCATCATCCAGGCCAGCCGACAGAGCGAGGAGGAGCTGGTGAACCGAG 2220  
 Db TACGCGCTGGGCATCATCCAGGCCAGCCGACAGAGCGAGGAGGAGCTGGTGAACCGAG 2220  
 Qy ATCATTCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGTGGCGCCGACCAAG 2280  
 Db ATCATTCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGTGGCGCCGACCAAG 2280  
 Qy GGATTCGGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
 Db GGATTCGGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
 Qy TTCTTGGACGCGATCGATGGCGGATCGTGATCTACCAAGTACATGAGCAGCTGTACGTTG 2400  
 Db TTCTTGGACGCGATCGATGGCGGATCGTGATCTACCAAGTACATGAGCAGCTGTACGTTG 2400  
 Qy GGACGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
 Db GGACGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457

## RESULT 2

ADM73766  
 ID ADM73766 standard; DNA; 2457 BP.  
 XX  
 AC ADM73766;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE HIV-1 polynucleotide #9.  
 XX  
 KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
 KW HIV type C protein; immunostimulant.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN US2003223961-A1.  
 XX  
 PD 04-DEC-2003.  
 XX  
 PF 05-JUL-2001; 2001US-00899575.  
 XX  
 PF 05-JUL-2000; 2000US-00610313.  
 PR  
 XX (MEGE/) MEGEDE J Z.  
 PA (BARN/) BARNETT S W.  
 PA (ENGE/) ENGELBRECHT S.  
 PA (RENS/) RENSBERG E J V.  
 XX  
 XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJ;

WPI; 2004-060515/06.

XX New expression cassette comprising a polynucleotide sequence encoding an  
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA  
 PT immunization, generating of packaging cell lines or in producing HIV Type  
 PT C proteins.

XX Claim 1; SEQ ID NO 32; 160pp; English.

XX The invention relates to an expression cassette comprising a  
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention

CC also relates to a recombinant expression system for use in a host cell  
 CC comprising an expression cassette, where the polynucleotide sequence  
 CC further comprises control elements capable of driving expression in the  
 CC selected host cell, a cell comprising an expression cassette where the  
 CC polynucleotide sequence further comprises control elements compatible  
 CC with the expression in the cell and a composition for generating an  
 CC immunological response, comprising an expression cassette. The expression  
 CC cassette and the methods of the invention are useful in eliciting an  
 CC immune response, in DNA immunisation, in generation of packaging cell  
 CC lines and in producing HIV Type C proteins. This sequence represents an  
 CC HIV-1 polynucleotide of the invention.

SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 100.0%; Score 2457; DB 12; Length 2457;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-296;  
 Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACGCCACCATGGCGAGGCCATGAGCAGGCCACCGCCAAACATCCTGATGCAG 60  
 Db 1 GTGACGCCACCATGGCGAGGCCATGAGCAGGCCACCGCCAAACATCCTGATGCAG 60  
 Qy 61 GCGAGCAACTTCAAGGGCCCCAGCGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
 Db 61 GCGAGCAACTTCAAGGGCCCCAGCGCATCATCAAGTCTTCAACTGGGCAAGAGGGC 120  
 Qy 121 CACATCCCGCAACTCCCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Db 121 CACATCCCGCAACTCCCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
 Qy 181 GGCACACAGATGAAGGACTGACCGAGCGCCAGCCAACTTCTTCGCGAGGACCTGGCC 240  
 Db 181 GGCACACAGATGAAGGACTGACCGAGCGCCAGCCAACTTCTTCGCGAGGACCTGGCC 240  
 Qy 241 TTCCCCCAGGCAAGGCCCGGAGTTCGCCAGCGAGCAGAACCCGGCCAAACAGCCCCACC 300  
 Db 241 TTCCCCCAGGCAAGGCCCGGAGTTCGCCAGCGAGCAGAACCCGGCCAAACAGCCCCACC 300  
 Qy 301 AGCGCGAGCTGCAGGTGCGCGCGCAACCCCCGAGCGAGCGCGCGCGAGCGCCAG 360  
 Db 301 AGCGCGAGCTGCAGGTGCGCGCGCAACCCCCGAGCGAGCGCGCGCGAGCGCCAG 360  
 Qy 361 GGCAACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGCATCAAGTG 420  
 Db 361 GGCAACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAAGCATCAAGTG 420  
 Qy 421 GCGGCGCAGATCAAGGAGGCGCTCTGAGACACCGCGCGCGAGCAGACCGTCTGGAGGAG 480  
 Db 421 GCGGCGCAGATCAAGGAGGCGCTCTGAGACACCGCGCGCGAGCAGACCGTCTGGAGGAG 480  
 Qy 481 ATGAGCCTGCCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540  
 Db 481 ATGAGCCTGCCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540  
 Qy 541 GTGCGCCAGTACGACACAGATCCTGATCGAGATCTGCGGCAAGAGAGGCATCGGCACCGTG 600  
 Db 541 GTGCGCCAGTACGACACAGATCCTGATCGAGATCTGCGGCAAGAGAGGCATCGGCACCGTG 600  
 Qy 601 CTGATCGGCGCCACCCCGGTGAACATCATCGCGCGCAACATGCTGAGCCAGCTGGGTGC 660  
 Db 601 CTGATCGGCGCCACCCCGGTGAACATCATCGCGCGCAACATGCTGAGCCAGCTGGGTGC 660  
 Qy 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG 720  
 Db 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATG 720  
 Qy 721 GAGCGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGAGGCCCTGACCGCC 780  
 Db 721 GAGCGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGAGGCCCTGACCGCC 780  
 Qy 781 ATTCGAGGAGATGGAGAGAGGGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTAC 840  
 Db 781 ATTCGAGGAGATGGAGAGAGGGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTAC 840

QY 841 AACACCCCGCTGTTCCGCATCAAGAAAGAGACACCAAGTGGCGCAAGCTGTGGAC 900  
Db 841 AACACCCCGCTGTTCCGCATCAAGAAAGAGACACCAAGTGGCGCAAGCTGTGGAC 900  
QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCGCCAC 960  
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCGCCAC 960  
QY 961 CCCGCGGCTGAAGAAGAAAGAGCGTGACCGTGTGGAGTGGGCGAGCGCTACTTTC 1020  
Db 961 CCCGCGGCTGAAGAAGAAAGAGCGTGACCGTGTGGAGTGGGCGAGCGCTACTTTC 1020  
QY 1021 AGCGTGCCCTGGAGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
Db 1021 AGCGTGCCCTGGAGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGCGATCCGCTACAGTACCAAGTGTGCGCCAGGGCTGGAAGGCGAGC 1140  
Db 1081 AACGAGACCCCGGCGATCCGCTACAGTACCAAGTGTGCGCCAGGGCTGGAAGGCGAGC 1140  
QY 1141 CCCAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGGCAACCCC 1200  
Db 1141 CCCAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGGCAACCCC 1200  
QY 1201 GAGATCGTGATCTACAGGCCCCCTGTAGTGGGCGAGCGCTGGAGATCGGCCAGCAC 1260  
Db 1201 GAGATCGTGATCTACAGGCCCCCTGTAGTGGGCGAGCGCTGGAGATCGGCCAGCAC 1260  
QY 1261 CGCGCAAGATCGAGGAGTGGCAAGCACCTGTGCGTGGGGCTTCCACCAACCCCGAC 1320  
Db 1261 CGCGCAAGATCGAGGAGTGGCAAGCACCTGTGCGTGGGGCTTCCACCAACCCCGAC 1320  
QY 1321 AAGAAGCACAGAGAGCCCCCTTCTGCCCATCGAGCTGCACCCCGACAAGTGGAGC 1380  
Db 1321 AAGAAGCACAGAGAGCCCCCTTCTGCCCATCGAGCTGCACCCCGACAAGTGGAGC 1380  
QY 1381 GTGCAGCCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAGAGCTG 1440  
Db 1381 GTGCAGCCCATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAGAGCTG 1440  
QY 1441 GTGGCAAGCTGAATGGGCCACGACAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTC 1500  
Db 1441 GTGGCAAGCTGAATGGGCCACGACAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTC 1500  
QY 1501 AAGCTGTGCGCGGCGCCAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGCGCGAG 1560  
Db 1501 AAGCTGTGCGCGGCGCCAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGCGCGAG 1560  
QY 1561 CTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGAGCCCGTSCACGGGTGTACTACGAC 1620  
Db 1561 CTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGAGCCCGTSCACGGGTGTACTACGAC 1620  
QY 1621 CCCAGCAAGGACCTGTGGCGCGAGATCCAGAGCAGGGCCACGACAGTGCACCTACCAG 1680  
Db 1621 CCCAGCAAGGACCTGTGGCGCGAGATCCAGAGCAGGGCCACGACAGTGCACCTACCAG 1680  
QY 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740  
Db 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1740  
QY 1741 CACACCAAGCAGTGAAGCAGTGAACCGAGCGGTGCAAGAGATCGCCATGGAGAGCATC 1800  
Db 1741 CACACCAAGCAGTGAAGCAGTGAACCGAGCGGTGCAAGAGATCGCCATGGAGAGCATC 1800  
QY 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGGAGACC 1860  
QY 1861 TGGTGACCGCATCTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTTGTGAACACCCCC 1920  
Db 1861 TGGTGACCGCATCTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTTGTGAACACCCCC 1920

QY 1921 CCCCTGTGAAGCTGTGTACAGCTGGAGAAAGAGCCCATCATCGCGCGCGAGACCTTC 1980  
Db 1921 CCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAAAGAGCCCATCATCGCGCGCGAGACCTTC 1980  
QY 1981 TACGTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTGACCGAC 2040  
Db 1981 TACGTGAGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGGCGGCTACGTGACCGAC 2040  
QY 2041 CGGGCGCGGCAAGATCGTGAAGCTGACCGAGACCAACCAACCGAGACCGAGCTGCAG 2100  
Db 2041 CGGGCGCGGCAAGATCGTGAAGCTGACCGAGACCAACCAACCGAGACCGAGCTGCAG 2100  
QY 2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGAGTGAACATCGTGACCGACCGCAG 2160  
Db 2101 GCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGCGAGTGAACATCGTGACCGACCGCAG 2160  
QY 2161 TACGCTTGGGCATCATCCAGGCCAGCCCCGACAAGAGCGAGCGAGCTGGTGAACCGAG 2220  
Db 2161 TACGCTTGGGCATCATCCAGGCCAGCCCCGACAAGAGCGAGCGAGCTGGTGAACCGAG 2220  
QY 2221 ATCATCGAGCAGCTGATCAAGAGGAAGGTGTACCTGAGCTGGTGCCGCCCCACAG 2280  
Db 2221 ATCATCGAGCAGCTGATCAAGAGGAAGGTGTACCTGAGCTGGTGCCGCCCCACAG 2280  
QY 2281 GGCAATCGGCGCAACGAGCAGATCGAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
Db 2281 GGCAATCGGCGCAACGAGCAGATCGAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTG 2340  
QY 2341 TTCTGTGACGCGATCGATGCGCGCATCTGATCTACCAAGTACATGAGACGACTGTACGCTG 2400  
Db 2341 TTCTGTGACGCGATCGATGCGCGCATCTGATCTACCAAGTACATGAGACGACTGTACGCTG 2400  
QY 2401 GGCAAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
Db 2401 GGCAAGCGGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
RESULT 3  
ACA03546  
ID ACA03546 standard; DNA; 2445 BP.  
XX ACA03546;  
XX AC  
XX AC  
DT 22-MAY-2003 (first entry)  
XX  
DE Synthetic DNA encoding immunogenic HIV peptide #29.  
XX  
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX  
OS Synthetic.  
XX  
XX WO2003004657-A1.  
XX  
XX 16-JAN-2003.  
XX  
XX 05-JUL-2002; 2002WO-US021421.  
XX  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
XX (CHIR ) CHIRON CORP.  
XX  
XX Zur Megede J, Barnett SW, Lian Y;  
XX  
XX WPI; 2003-221602/21.  
DR  
XX  
XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
XX polypeptides, useful as immunogenic compositions or vaccines for  
PT



generating humoral or cellular immune responses against HIV in a subject, especially humans.

Example 1; Fig 34; 262pp; English.

The invention describes a synthetic polynucleotide encoding 2 or more immunogenic HIV polypeptides, where at least 2 of the polypeptides are derived from different HIV subtypes. The polynucleotide is useful for immunisation, generation of packaging cell lines, or production of HIV polypeptides. The polynucleotide and its encoded proteins are useful as immunogenic compositions or vaccines for generating humoral or cellular immune responses against HIV in a subject, or for inducing neutralising antibodies against HIV. The gene delivery vector comprising the polynucleotide is also useful for DNA immunisation of, or for generating an immune response (e.g. a humoral or cellular immune response) in, a subject such as a mammal, particularly a human. This sequence encodes a human immunodeficiency virus immunogenic peptide

Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 8; Length 2445;

Best Local Similarity 100.0%; Pred. No. 4.7e-294;

Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	7	GCCACCATGGCCGAGGCCATGAGCCAGCCACACAGCGCCAAACATCTCTGATGACGCGCAGC	66
Db	1	GCCACCATGGCCGAGGCCATGAGCCAGCCACACAGCGCCAAACATCTCTGATGACGCGCAGC	60
Qy	67	AACCTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATC	126
Db	61	AACCTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATC	120
Qy	127	GCCCGCAACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGGCCAC	186
Db	121	GCCCGCAACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGGCCAC	180
Qy	187	CAGATGAAGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCC	246
Db	181	CAGATGAAGGACTGCACCGAGCGCCAGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCC	240
Qy	247	CAGGGCAAGGCCCGCGAGTTTCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGC	306
Db	241	CAGGGCAAGGCCCGCGAGTTTCCAGCGAGCAGAAACCGCGCCAAACAGCCCCACAGCGGC	300
Qy	307	GAGCTGCAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCAC	366
Db	301	GAGCTGCAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCAC	360
Qy	367	CTGAACTTCCCCCAGATCACCTTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC	426
Db	361	CTGAACTTCCCCCAGATCACCTTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC	420
Qy	427	CAGATCAAGGAGGCCCTGTGGACACCGCGCGCCGACGACACCGTGTGGAGGATGAGC	486
Db	421	CAGATCAAGGAGGCCCTGTGGACACCGCGCGCCGACGACACCGTGTGGAGGATGAGC	480
Qy	487	CTGCCCGCGAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGGC	546
Db	481	CTGCCCGCGAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGGC	540
Qy	547	CAGTACGACCAAGATCTCTGATCGAGATCTGCGCGCAAGAGGCCATCGCGCATCGTCTGATC	606
Db	541	CAGTACGACCAAGATCTCTGATCGAGATCTGCGCGCAAGAGGCCATCGCGCATCGTCTGATC	600
Qy	607	GGCCCCACCCCGTGAAACATCATCGCGCGCAACATGTGTACCCAGCTGGGCTGACCCCTG	666
Db	601	GGCCCCACCCCGTGAAACATCATCGCGCGCAACATGTGTACCCAGCTGGGCTGACCCCTG	660
Qy	667	AACCTTCCCCATCAGCCCCATCGACCGTGCCTGCTGAGCTGAGCCCGGCGATGACGCGC	726
Db	661	AACCTTCCCCATCAGCCCCATCGACCGTGCCTGCTGAGCTGAGCCCGGCGATGACGCGC	720
Qy	727	CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGGCCCTGACCGCCATCTGC	786

Db	721	CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGGCCCTGACCGCCATCTGC	780
Qy	787	GAGGAGATGAGAGGAGGGCCAGATCACCAGATCGGCCCGGAGNACCCCTACAAACACC	846
Db	781	GAGGAGATGAGAGGAGGGCCAGATCACCAGATCGGCCCGGAGNACCCCTACAAACACC	840
Qy	847	CCCGTGTTCGCCATCAAGAAGAAGACAGACCAAGTGGCGCAAGCTGGTGGACTTCCGC	906
Db	841	CCCGTGTTCGCCATCAAGAAGAAGACAGACCAAGTGGCGCAAGCTGGTGGACTTCCGC	900
Qy	907	GAGCTGAACCAAGGCCACCCAGGACTTCTGGAGAGTGAGCTGGGCATCCCCACCCGCC	966
Db	901	GAGCTGAACCAAGGCCACCCAGGACTTCTGGAGAGTGAGCTGGGCATCCCCACCCGCC	960
Qy	967	GGCTTGAAGAGAGAGAGCGGTGACCGTGTGACGTGGCGGACGCTACTTTCAGGGTG	1026
Db	961	GGCTTGAAGAGAGAGAGCGGTGACCGTGTGACGTGGCGGACGCTACTTTCAGGGTG	1020
Qy	1027	CCCTGGAAGGAGACTTCCGCAAGTACACCGCCCTTCAACCATCCCCAGCATCAACAAAGAG	1086
Db	1021	CCCTGGAAGGAGACTTCCGCAAGTACACCGCCCTTCAACCATCCCCAGCATCAACAAAGAG	1080
Qy	1087	ACCCCGGCAATCCGCTACCAAGTACAAAGTGTGCCCCAGGGCTGGAGGGCAGCCCCCAGC	1146
Db	1081	ACCCCGGCAATCCGCTACCAAGTACAAAGTGTGCCCCAGGGCTGGAGGGCAGCCCCCAGC	1140
Qy	1147	ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGGCGCGGCAACCCCGAGATC	1206
Db	1141	ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGGCGCGGCAACCCCGAGATC	1200
Qy	1207	GTGATCTTACCAAGGCCCCCTGTACGTGGCAGCAGCCTGGAGATCGGCACACCGCGCC	1266
Db	1201	GTGATCTTACCAAGGCCCCCTGTACGTGGCAGCAGCCTGGAGATCGGCACACCGCGCC	1260
Qy	1267	AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCGCAAGAAG	1326
Db	1261	AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCGCAAGAAG	1320
Qy	1327	CACCAAGAGAGCCCCCTTCTTCCCATCGAGCTGCAACCCCGACAAAGTGAACGCTGAG	1386
Db	1321	CACCAAGAGAGCCCCCTTCTTCCCATCGAGCTGCAACCCCGACAAAGTGAACGCTGAG	1380
Qy	1387	CCCATCGAGCTGCCCGAGAGAGGAGCTGGAACCGTGAACGACATCCAGNAGCTTGTGGC	1446
Db	1381	CCCATCGAGCTGCCCGAGAGAGGAGCTGGAACCGTGAACGACATCCAGNAGCTTGTGGC	1440
Qy	1447	AAGCTGAATGGGCCAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG	1506
Db	1441	AAGCTGAATGGGCCAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG	1500
Qy	1507	CTGGCGGCGCCAAAGGCCCTGACCGACATCGTGGCCCTGACCGAGGAGGCCGAGCTGAG	1566
Db	1501	CTGGCGGCGCCAAAGGCCCTGACCGACATCGTGGCCCTGACCGAGGAGGCCGAGCTGAG	1560
Qy	1567	CTGGCCGAGAACCGCGAGATCTTGGCGGAGCCCGTGCACGCGCTGTACTACAGCCCGAGC	1626
Db	1561	CTGGCCGAGAACCGCGAGATCTTGGCGGAGCCCGTGCACGCGCTGTACTACAGCCCGAGC	1620
Qy	1627	AAGSACCTGGTGGCCGAGATCCAGAAGCAGGGGCCACGACCAAGTGGACCTTACAGATCTAC	1686
Db	1621	AAGSACCTGGTGGCCGAGATCCAGAAGCAGGGGCCACGACCAAGTGGACCTTACAGATCTAC	1680
Qy	1687	CAGGAGCCCTTCAAGAACTCTGAAGACCGGCAGTACGCGCAAGATGCGCACCGGCCACACC	1746
Db	1681	CAGGAGCCCTTCAAGAACTCTGAAGACCGGCAGTACGCGCAAGATGCGCACCGGCCACACC	1740
Qy	1747	AACGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAGAGCATCGTGATC	1806
Db	1741	AACGAGCTGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAGAGCATCGTGATC	1800
Qy	1807	TGGGGCAAGACCCCAAGTTCCGCTTCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGG	1866



Db 661 AACTTCCCATCAGCCCATCGAGACCGTCCCGCTGAAGCTGAAGCCCGGCGATCGACGGC 720  
Qy 727 CCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 786  
Db 721 CCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAGATCAAGGGCCCTGACCGCCATCTGC 780  
Qy 787 GAGGAGATGGAAGAGAGGGGCAAGATCAACAGATCGGCCCGGAGAACCCCTCAACAACC 846  
Db 781 GAGGAGATGGAAGAGAGGGGCAAGATCAACAGATCGGCCCGGAGAACCCCTCAACAACC 840  
Qy 847 CCGCTGTTCCGCATCAAGAGAAGGAGCAGCACCAGTGGCGCAAGCTGGTGACTTCCGC 906  
Db 841 CCGCTGTTCCGCATCAAGAGAAGGAGCAGCACCAGTGGCGCAAGCTGGTGACTTCCGC 900  
Qy 907 GAGCTGAACAAAGCGCACCAGGACTTCTGGAGAGTGCAGCTGGGCATCCCCCACC CGCC 966  
Db 901 GAGCTGAACAAAGCGCACCAGGACTTCTGGAGAGTGCAGCTGGGCATCCCCCACC CGCC 960  
Qy 967 GGCCTGAAGAGAAGAGAGCGTGACCGTGTGACAGTGGGCGAGCGCTACTTTCAGCGTG 1026  
Db 961 GGCCTGAAGAGAAGAGAGCGTGACCGTGTGACAGTGGGCGAGCGCTACTTTCAGCGTG 1020  
Qy 1027 CCGCTGAGCAGGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAGAG 1086  
Db 1021 CCGCTGAGCAGGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAGAG 1080  
Qy 1087 ACCCCCGGCATCCCGTACAGTACAAAGTGTGCTGCCAGGGCTGGAAGGGCAGCGCCAGC 1146  
Db 1081 ACCCCCGGCATCCCGTACAGTACAAAGTGTGCTGCCAGGGCTGGAAGGGCAGCGCCAGC 1140  
Qy 1147 ATCTTCAGAGCAGATGACCAAGATCTGAGCGCTTCCGCGCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCAGAGCAGATGACCAAGATCTGAGCGCTTCCGCGCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACCAAGCGCCCTGTAGCTGGGAGCGACCTGGAGATCGGCGACCGCGCC 1266  
Db 1201 GTGATCTACCAAGCGCCCTGTAGCTGGGAGCGACCTGGAGATCGGCGACCGCGCC 1260  
Qy 1267 AAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGGCTTCAACACCCCGCAAGAAG 1326  
Db 1261 AAGATCGAGGAGCTGCGCAAGCAGCTGTGCGCTGGGGCTTCAACACCCCGCAAGAAG 1320  
Qy 1327 CACCAAGAGGAGCGCCCTTCTGCCATCGAGCTGACCCCGCAAGTGGACCGTGCAG 1386  
Db 1321 CACCAAGAGGAGCGCCCTTCTGCCATCGAGCTGACCCCGCAAGTGGACCGTGCAG 1380  
Qy 1387 CCCATCGAGCTGCCGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGC 1446  
Db 1381 CCCATCGAGCTGCCGAGGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGC 1440  
Qy 1447 AAGCTGAATCGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1506  
Db 1441 AAGCTGAATCGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTG 1500  
Qy 1507 CTGCGCGCGCCAAAGGCGCTGACCGACATGTGCCCCCTGACCGAGGAGCGCGAGCTGGAG 1566  
Db 1501 CTGCGCGCGCCAAAGGCGCTGACCGACATGTGCCCCCTGACCGAGGAGCGCGAGCTGGAG 1560  
Qy 1567 CTGCGCGAGAACCGGAGATCTCTGGGAGCGCCGTGACCGCGTGTACTACGACCCCGAGC 1626  
Db 1561 CTGCGCGAGAACCGGAGATCTCTGGGAGCGCCGTGACCGCGTGTACTACGACCCCGAGC 1620  
Qy 1627 AAGGACCTGTGGCGGAGATCCAGAAGCAGGGCCACGACAGTGGACCTACAGATCTAC 1686  
Db 1621 AAGGACCTGTGGCGGAGATCCAGAAGCAGGGCCACGACAGTGGACCTACAGATCTAC 1680  
Qy 1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCGCACCGCCACACC 1746  
Db 1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCGCACCGCCACACC 1740  
Qy 1747 AACGACGTGAAGCAGTGCACGAGCGCGTGACAGAGATCGCCATGGAGAGATCGTGATC 1806  
Db 1741 AACGACGTGAAGCAGTGCACGAGCGCGTGACAGAGATCGCCATGGAGAGATCGTGATC 1800

Qy 1807 TGGGCAAGACCCCACAGTTCCGCTCCCATCCAGAGGAGACCTGGAGACCTGTGTGG 1866  
Db 1801 TGGGCAAGACCCCACAGTTCCGCTCCCATCCAGAGGAGACCTGGAGACCTGTGTGG 1860  
Qy 1867 ACCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCTG 1926  
Db 1861 ACCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCTG 1920  
Qy 1927 GTGAAGCTGTGTACCAAGCTGGAGAGGAGCCCATCATCGCGCGGAGACCTTCTACGTG 1986  
Db 1921 GTGAAGCTGTGTACCAAGCTGGAGAGGAGCCCATCATCGCGCGGAGACCTTCTACGTG 1980  
Qy 1987 GACGCGCCCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACTGTGAACCGAGCGGGC 2046  
Db 1981 GACGCGCCCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACTGTGAACCGAGCGGGC 2040  
Qy 2047 CGGCAAGATCTGTGAGCTTGAAGCCGAGACCAACCAAGAGACCGAGCTGCAGGCCATC 2106  
Db 2041 CGGCAAGATCTGTGAGCTTGAAGCCGAGACCAACCAAGAGACCGAGCTGCAGGCCATC 2100  
Qy 2107 CAGCTGCCCTTGCAGGACAGCGCAGCAGAGTGAACATCGTGACCGACAGCAGTACGCC 2166  
Db 2101 CAGCTGCCCTTGCAGGACAGCGCAGCAGAGTGAACATCGTGACCGACAGCAGTACGCC 2160  
Qy 2167 CTGGGCATCATCCAGGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCGAGATCATC 2226  
Db 2161 CTGGGCATCATCCAGGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCGAGATCATC 2220  
Qy 2227 GAGCAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGGTGCCCGGCCCAAGAGGCCATC 2286  
Db 2221 GAGCAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGGTGCCCGGCCCAAGAGGCCATC 2280  
Qy 2287 GGGGCAACGAGCAGATCGACAGCTGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCTG 2346  
Db 2281 GGGGCAACGAGCAGATCGACAGCTGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCTG 2340  
Qy 2347 GACGCGATCGATGGCGGCTGTGATCTACAGTACATGACGACCTGTGTGGGCGAGC 2406  
Db 2341 GACGCGATCGATGGCGGCTGTGATCTACAGTACATGACGACCTGTGTGGGCGAGC 2400  
Qy 2407 GGGGCGCTTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 2401 GGGGCGCTTAGGATCGATTAAGCTTCCCGGGCTAGCACCGGT 2445

## RESULT 5

ADCL3230  
ID ADC13230 standard; DNA; 3930 BP.

XX AC ADC13230;

XX DT 18-DEC-2003 (first entry)

XX DNA of HIV construct GagCompPolmut\_C SEQ ID NO 9.

XX DE expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
XX KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX OS Human immunodeficiency virus.

XX PN WO2003004620-A2.

XX PD 16-JAN-2003.

XX PF 05-JUL-2002; 2002WO-US021420.

XX PR 05-JUL-2001; 2001US-0303192P.

XX PR 31-AUG-2001; 2001US-0316860P.

XX PR 16-JAN-2002; 2002US-0349871P.

XX PA (CHIR ) CHIRON CORP.

PA (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 PI WPI; 2003-221593/21.  
 XX  
 PT New expression cassette comprising a polynucleotide sequence encoding a  
 PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 PT Prot, or Rev polypeptide, useful for immunization, or generating  
 PT packaging cell lines.  
 XX  
 PS Disclosure; Fig 6; 301pp; English.  
 XX  
 CC The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intramuscularly, intravenously,  
 CC subcutaneously, intradermally, transdermally, intravaginally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX  
 SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.4; DB 10; Length 3930;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-293;  
 Matches 2437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 14 TGGCGAGGCGCATGAGCGGCGCAGCGCCAGCATCTGATGAGCGCGCACTTCA 73  
 DB 1487 TGGCGAGGCGCATGAGCGGCGCAGCGCCAGCATCTGATGAGCGCGCACTTCA 1546

QY 74 AGGGCCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGCGGCACATCGCCGCA 133  
 DB 1547 AGGGCCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGCGGCACATCGCCGCA 1606

QY 134 ACTGCGGCGCCCCGCGCAAGAGGCGTGTGAGTGGCGGCAAGAGGCGGCACAGATGA 193  
 DB 1607 ACTGCGGCGCCCCGCGCAAGAGGCGTGTGAGTGGCGGCAAGAGGCGGCACAGATGA 1666

QY 194 AGGACTGCACCGAGCGCGCAGGCGCAACTTCTCCGAGGAGCTGCGCTTCCCGCAGGCA 253  
 DB 1667 AGGACTGCACCGAGCGCGCAGGCGCAACTTCTCCGAGGAGCTGCGCTTCCCGCAGGCA 1726

QY 254 AGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCCAAAGCGCCCAAGCGCGAGCTGC 313  
 DB 1727 AGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCCAAAGCGCCCAAGCGCGAGCTGC 1786

QY 314 AGGTGCGGCGGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 DB 1787 AGGTGCGGCGGCGCAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846

QY 374 TCCCGCAGATCACTGTGGAGCGCCCTCTGAGGATCAAGTGTGGCGCGCGAGATCA 433  
 DB 1847 TCCCGCAGATCACTGTGGAGCGCCCTCTGAGGATCAAGTGTGGCGCGCGAGATCA 1906

QY 434 AGGAGGCGCTGTGGACACCGCGCGCGAGCACCGCTGCTGGAGGAGATGAGCTGCGCG 493  
 DB 1907 AGGAGGCGCTGTGGACACCGCGCGCGAGCACCGCTGCTGGAGGAGATGAGCTGCGCG 1966

QY 494 GCAAGTGGAGCGCGAGATGATCGGCGGCGATCGGCGGCTTCAATCAAGTGGCGCGAGTAG 553  
 DB 1967 GCAAGTGGAGCGCGAGATGATCGGCGGCGATCGGCGGCTTCAATCAAGTGGCGCGAGTAG 2026

QY 554 ACCAGATCTGATCGAGATCTCGCGCAAGAGCGCATCGGCGCTGTGATCGCGCGCA 613  
 DB 2027 ACCAGATCTGATCGAGATCTCGCGCAAGAGCGCATCGGCGCTGTGATCGCGCGCA 2086

QY 614 CCCCCTGAAACATATCGGCGCGCAATGCTGACCGAGCTGGGCTGCACTTCACTTCC 673  
 DB 2087 CCCCCTGAAACATATCGGCGCGCAATGCTGACCGAGCTGGGCTGCACTTCACTTCC 2146

QY 674 CCATCAGCCCCCATCGAGACCGTGGCCGCTGAGCTGAAGCCCGGCATGACGCGCCCCAAGG 733  
 DB 2147 CCATCAGCCCCCATCGAGACCGTGGCCGCTGAGCTGAAGCCCGGCATGACGCGCCCCAAGG 2206

QY 734 TGAAGCAGTGGCCCCCTGACCCGAGGAGAGATCAAGGCCCTTGAACGCCCATCTGCGAGGAGA 793  
 DB 2207 TGAAGCAGTGGCCCCCTGACCCGAGGAGAGATCAAGGCCCTTGAACGCCCATCTGCGAGGAGA 2266

QY 794 TGGAGAGGAGGCGCAAGATCAACAGATCGCGCCCGAGAACCCCTTACAAACCCCGCTGT 853  
 DB 2267 TGGAGAGGAGGCGCAAGATCAACAGATCGCGCCCGAGAACCCCTTACAAACCCCGCTGT 2326

QY 854 TCGCCATCAAG 913  
 DB 2327 TCGCCATCAAG 2386

QY 914 ACAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGGATCCCCACCCCGCGGCTTGA 973  
 DB 2387 ACAAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGGATCCCCACCCCGCGGCTTGA 2446

QY 974 AGAAG 1033  
 DB 2447 AGAAG 2506

QY 1034 ACGAGGACTTCGCGCAAGTACACCGCTTCAACATCCCGCAGCATCAACAGAGACCCCG 1093  
 DB 2507 ACGAGGACTTCGCGCAAGTACACCGCTTCAACATCCCGCAGCATCAACAGAGACCCCG 2566

QY 1094 GCATCCGCTACAGTACAAACGCTGTGCGCCAGGCTGGAAGGCGAGCGCCAGCATCTTCC 1153  
 DB 2567 GCATCCGCTACAGTACAAACGCTGTGCGCCAGGCTGGAAGGCGAGCGCCAGCATCTTCC 2626

QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCTTCCGCGCCCGCAACCCCGAGATCTGATCT 1213  
 DB 2627 AGAGCAGCATGACCAAGATCTTGGAGCTTCCGCGCCCGCAACCCCGAGATCTGATCT 2686

QY 1214 ACCAGGCCCCCTGTACGTGGGCGAGGAGCTGAGAGTGGCGCAGCACCGCGCAGAGATCG 1273  
 DB 2687 ACCAGGCCCCCTGTACGTGGGCGAGGAGCTGAGAGTGGCGCAGCACCGCGCAGAGATCG 2746

QY 1274 AGGAGCTCGCAAGCACTGTGCGCTGGGCTTCCACACCCCGCAGAGAGAGAGAGAGAG 1333  
 DB 2747 AGGAGCTCGCAAGCACTGTGCGCTGGGCTTCCACACCCCGCAGAGAGAGAGAGAGAG 2806

QY 1334 AGGAGCCCCCTTCTGCGCATCGAGCTGCAACCGCGCAAGTGGAGCGGTGAGAGAGAGAG 1393  
 DB 2807 AGGAGCCCCCTTCTGCGCATCGAGCTGCAACCGCGCAAGTGGAGCGGTGAGAGAGAGAG 2866

QY 1394 AGCTGCGCGAG 1453  
 DB 2867 AGCTGCGCGAG 2926

QY 1454 ACTGGGCGAG 1513  
 DB 2927 ACTGGGCGAG 2986

QY 1514 GCGCGAAGGCGCTGACCGAGATCTGCGCCCTTGAACCGAGAGAGAGAGAGAGAGAGAGAG 1573  
 DB 2987 GCGCGAAGGCGCTGACCGAGATCTGCGCCCTTGAACCGAGAGAGAGAGAGAGAGAGAGAG 3046

QY 1574 AGAACCGGAGAGATCTGCGCGAGGCGCGGTGACCGGCTGTACTACGACCCCGAGAGAGAG 1633  
 DB 3047 AGAACCGGAGAGATCTGCGCGAGGCGCGGTGACCGGCTGTACTACGACCCCGAGAGAGAG 3106

QY 1634 TGGTGGCGGAGATCCAG 1693  
 DB 3107 TGGTGGCGGAGATCCAG 3166

QY 1694 CTTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATGCGGCAACCGCGCGAGAGAGAG 1753  
 DB 3167 CTTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATGCGGCAACCGCGCGAGAGAGAG 3226



Db 601 |||||CTGATCGGCCCAACCCCGTGAACATCATCGCGCGAACAATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAACTTCCCATCAGCCCCATCAGACCCGTCGCGTGAAGCTGAAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCCATCAGACCCGTCGCGTGAAGCTGAAAGCCCGGCATG 720  
QY 721 GACGCGCCCAAGTGAACAGTGGCCCTGACCGGAGGAGATCAAGGCTGACCGCC 780  
Db 721 GACGCGCCCAAGTGAACAGTGGCCCTGACCGGAGGAGATCAAGGCTGACCGCC 780  
QY 781 ATCTCGAGGAGATGGAAGAGAGGCGCAAGATCACCAAGATCGGCCCGAGAACCCCTAC 840  
Db 781 ATCTCGAGGAGATGGAAGAGAGGCGCAAGATCACCAAGATCGGCCCGAGAACCCCTAC 840  
QY 841 AACACCCCGTGTTCGCCATCAAGAAGAGAGACACCAAGTGGCGCAAGCTGTGTGAC 900  
Db 841 AACACCCCGTGTTCGCCATCAAGAAGAGAGACACCAAGTGGCGCAAGCTGTGTGAC 900  
QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGCATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGCATCCCCAC 960  
QY 961 CCGCGCGCTGAAGAAGAGAGCGTGAACCGTCTGGAGCGTGGCGGACGCGCTACTTC 1020  
Db 961 CCGCGCGCTGAAGAAGAGAGCGTGAACCGTCTGGAGCGTGGCGGACGCGCTACTTC 1020  
QY 1021 AGCGTCCCTGGACGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080  
Db 1021 AGCGTCCCTGGACGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGATCGCTACAGTACAGTCTGCTGCGGCGTGGCGGCTGGAGGGGAGC 1140  
Db 1081 AACGAGACCCCGGATCGCTACAGTACAGTCTGCTGCGGCGTGGCGGCTGGAGGGGAGC 1140  
QY 1141 CCCAGCATCTTCAGAGCAGCATCAACAGATCTCGAGCGCTTCCGCGCGCGCAACCCC 1200  
Db 1141 CCCAGCATCTTCAGAGCAGCATCAACAGATCTCGAGCGCTTCCGCGCGCGCAACCCC 1200  
QY 1201 GAGATCGTGATCTACAGGCCCTTCTGATCGGCGAGCGACTCGAGATCGGCGAGCAGC 1260  
Db 1201 GAGATCGTGATCTACAGGCCCTTCTGATCGGCGAGCGACTCGAGATCGGCGAGCAGC 1260  
QY 1261 CCGCGCAAGATCGAGAGCTGCGCAAGCACCTGCTGCGTGGGCTTCAACACCCCGGAC 1320  
Db 1261 CCGCGCAAGATCGAGAGCTGCGCAAGCACCTGCTGCGTGGGCTTCAACACCCCGGAC 1320  
QY 1321 AAGAAGCACCAAGAGAGCCCTTCTGCGCCAT-----CGAGCTGCACCCCGACAG 1374  
Db 1321 AAGAAGCACCAAGAGAGCCCTTCTGCGCCATCGAGTGGGCTAGAGCTGCACCCCGACAG 1380  
QY 1375 TGGACCGTGCAGCCCATCAGATGCGCGAGAGAGAGTGAACCGTGAACGACATCCAG 1434  
Db 1381 TGGACCGTGCAGCCCATCAGATGCGCGAGAGAGAGTGAACCGTGAACGACATCCAG 1440  
QY 1435 AGCTGGTGGGCAAGCTGACCTGGGCGAGCGAGATCTACCCCGGATCAAGTGGCGGAG 1494  
Db 1441 AAGCTGGTGGGCAAGCTGACCTGGGCGAGCGAGATCTACCCCGGATCAAGTGGCGGAG 1500  
QY 1495 CTGTGCAAGCTGCTGCGGCGCCCAAGGCCCTGACCGACATCGTGCCTTACCCGAGGAG 1554  
Db 1501 CTGTGCAAGCTGCTGCGGCGCCCAAGGCCCTGACCGACATCGTGCCTTACCCGAGGAG 1560  
QY 1555 GCCAGCTGGAGCTGGCGAGAACCCGAGATCTCTGCGGAGCGCCCGTGCACCGCTGTAC 1614  
Db 1561 GCCAGCTGGAGCTGGCGAGAACCCGAGATCTCTGCGGAGCGCCCGTGCACCGCTGTAC 1620  
QY 1615 TACGACCCCGACAGACTGGTGGCGGAGTCCAGAGACGAGGCGACGACAGTGGACC 1674  
Db 1621 TACGACCCCGACAGACTGGTGGCGGAGTCCAGAGACGAGGCGACGACAGTGGACC 1680  
QY 1675 TACCAGATCTACAGAGCGCTTCAAGAACCTTCAAGACCGGCAAGTACGCCAAGATGGC 1734

Db 1681 TACCAGATCTACAGAGCGCTTCAAGAACCTTCAAGAACCGGCAAGTACGCCAAGATGGC 1740  
QY 1735 ACCGCCCHACACACAGAGCTGAAGAGCTGACCGAGCGCGTGACAGAAATGCGCCATGAG 1794  
Db 1741 ACCGCCCHACACACAGAGCTGAAGAGCTGACCGAGCGCGTGACAGAAATGCGCCATGAG 1800  
QY 1795 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAAAGAGACCTGG 1854  
Db 1801 AGCATCGTGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAAAGAGACCTGG 1860  
QY 1855 GAGACCTGGTGGACCGACTACTTGGCAGGCCACCTGGATGCCCGAGTGGGAGTTCGTGAAC 1914  
Db 1861 GAGACCTGGTGGACCGACTACTTGGCAGGCCACCTGGATGCCCGAGTGGGAGTTCGTGAAC 1920  
QY 1915 ACCCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGCGCGGAG 1974  
Db 1921 ACCCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGCGCGGAG 1980  
QY 1975 ACCTTCTACGTGGACGGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGCTG 2034  
Db 1981 ACCTTCTACGTGGACGGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGCTG 2040  
QY 2035 ACCGACCGGGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGAGACCGAG 2094  
Db 2041 ACCGACCGGGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAG 2100  
QY 2095 CTGACGGCCATCTCAGCTGGGCCCTGAGSACAGCGCAGCGAGGTGAACATCGTGAACGAC 2154  
Db 2101 CTGACGGCCATCTCAGCTGGGCCCTGAGSACAGCGCAGCGAGGTGAACATCGTGAACGAC 2160  
QY 2155 AGCAGTACGCGCTGGGCGATCATCAGGCCCGAGCCGACAAAGAGCGAGCGAGCTGGTG 2214  
Db 2161 AGCAGTACGCGCTGGGCGATCATCAGGCCCGAGCCGACAAAGAGCGAGCGAGCTGGTG 2220  
QY 2215 AACCAGATCATCAGCAGCTGATCAAGAGAGAGAGTGTACTCTGAGCTGGTGGCCGCC 2274  
Db 2221 AACCAGATCATCAGCAGCTGATCAAGAGAGAGAGTGTACTCTGAGCTGGTGGCCGCC 2280  
QY 2275 CACAAGGCGATCGGCGCAACGAGCAGATCAAGAGCTGGTGAAGAGGCGATCCGCAAG 2334  
Db 2281 CACAAGGCGATCGGCGCAACGAGCAGATCAAGAGCTGGTGAAGAGGCGATCCGCAAG 2340  
QY 2335 GTGCTGTTCTTGGAGCGGCGATCGAGCGGCGATCGTGAATCTACAGTACATGGACGACCTG 2394  
Db 2341 GTGCTGTTCTTGGAGCGGCGATCGAGCGGCGATCGTGAATCTACAGTACATGGACGACCTG 2400  
QY 2395 TACGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2454  
Db 2401 TACGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
QY 2455 TTC 2457  
Db 2461 TTC 2463

## RESULT 7

ADM73765

ID ADM73765 standard; DNA; 2463 BP.

XX

AC ADM73765;

XX

DT 03-JUN-2004 (first entry)

XX

DE HIV-1 polynucleotide #8.

XX

KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;

KW HIV type C protein; immunostimulant.

XX

OS Human immunodeficiency virus 1.

XX

PN US2003223961-A1.

XX

PD 04-DEC-2003.



XX	05-JUL-2001; 2001US-00899575.	Qy	481	ATGAGCCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG	540
PF	XX	Db	481	ATGAGCCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG	540
PR	05-JUL-2000; 2000US-00610313.	Qy	541	GTGCGCCAGTACGACAGATCCTGATCGAGATCTGCGCGCAGAGAGCCATCGGCACCGTG	600
PA	(MEGE/) MEGEDE J Z.	Db	541	GTGCGCCAGTACGACAGATCCTGATCGAGATCTGCGCGCAGAGAGCCATCGGCACCGTG	600
PA	(BARN/) BARNETT S W.	Qy	601	CTGATCGCGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC	660
PA	(ENGE/) ENGELBRECHT S.	Db	601	CTGATCGCGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC	660
XX	(RENS/) RENSBERG E J V.	Qy	661	ACCTGAACTTCCCATCAGCCCCCATCGAGACCGTGCCTGAAAGTGAAGCCCGGATG	720
PI	Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJ;	Db	661	ACCTGAACTTCCCATCAGCCCCCATCGAGACCGTGCCTGAAAGTGAAGCCCGGATG	720
XX	WPI: 2004-060515/06.	Qy	721	GACGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC	780
XX	New expression cassette comprising a polynucleotide sequence encoding an	Db	721	GACGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC	780
XX	HIV Pol polypeptide, useful in eliciting an immune response, in DNA	Qy	781	ATCTGCGAGGAGTGGAGAGAGGGCAAGATCACCAAGATCGCGCCCGCGAAGCCCTTAC	840
XX	immunization, generating of packaging cell lines or in producing HIV Type	Db	781	ATCTGCGAGGAGTGGAGAGAGGGCAAGATCACCAAGATCGCGCCCGCGAAGCCCTTAC	840
XX	C proteins.	Qy	841	AACACCCCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900
PS	Claim 1; SEQ ID NO 31; 160pp; English.	Db	841	AACACCCCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900
XX	The invention relates to an expression cassette comprising a	Qy	901	TTCCGCGAGTGAACAG	960
CC	polynucleotide sequence encoding an HIV Pol polypeptide. The invention	Db	901	TTCCGCGAGTGAACAG	960
CC	also relates to a recombinant expression system for use in a host cell	Qy	961	CCCGCCCGCTGAAAG	1020
CC	comprising an expression cassette, where the polynucleotide sequence	Db	961	CCCGCCCGCTGAAAG	1020
CC	further comprises control elements capable of driving expression in the	Qy	1021	AGCGTGCCTTGAAG	1080
CC	selected host cell, a cell comprising an expression cassette where the	Db	1021	AGCGTGCCTTGAAG	1080
CC	polynucleotide sequence further comprises control elements compatible	Qy	1081	AAAG	1140
CC	with the expression in the cell and a composition for generating an	Db	1081	AAAG	1140
CC	immunological response, comprising an expression cassette. The expression	Qy	1141	CCAGAGATCTTCCAG	1200
CC	cassette and the methods of the invention are useful in eliciting an	Db	1141	CCAGAGATCTTCCAG	1200
CC	immune response, in DNA immunisation, in generation of packaging cell	Qy	1201	GAGATCGTATCTACCAAGGCCCCCTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1260
CC	lines and in producing HIV Type C proteins. This sequence represents an	Db	1201	GAGATCGTATCTACCAAGGCCCCCTTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1260
CC	HIV-1 polynucleotide of the invention.	Qy	1261	CGCGCCAGAGATCGAG	1320
XX	Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;	Db	1261	CGCGCCAGAGATCGAG	1320
XX	Query Match 99.2%; Score 2436.2; DB 12; Length 2463;	Qy	1321	AAGAGACACAG	1374
XX	Best Local Similarity 99.6%; Pred. No. 3.6e-293;	Db	1321	AAGAGACACAG	1374
XX	Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;	Qy	1375	TGACCCGTGACGCCCCATCGAGTGCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1434
Qy	1 GTGACGCCACCATGCGCGAGGCCATGAGCAGCGCACCGCCACCAATCTCTGATGACG	Db	1381	TGACCCGTGACGCCCCATCGAGTGCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1440
Db	1 GTGACGCCACCATGCGCGAGGCCATGAGCAGCGCACCGCCACCAATCTCTGATGACG	Qy	1435	AACTGTGTGGGAGAGTGAATCTGGCCAGCAGATCTACCCCGGCATCAAGGTGCGCGAG	1494
Qy	61 CGCAGCAACTTCAAGGSCCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGAGGSC	Db	1441	AACTGTGTGGGAGAGTGAATCTGGCCAGCAGATCTACCCCGGCATCAAGGTGCGCGAG	1494
Db	61 CGCAGCAACTTCAAGGSCCCCAAGCGCATCATCAAGTGTCTCAACTCGCGCAAGAGGSC	Qy	1495	CTGTGCAAGTGTGTGCGCGCGCCCAAGCGCTTACCGAGCATGTGCGCCCTGACCGAGAG	1554
Qy	121 CACATCGCCCGCAACTGCGCGGCCCGCGCAAGAGGGCTGCTGGAAGTGGGCAAGGAG	Db	1501	CTGTGCAAGTGTGTGCGCGCGCCCAAGCGCTTACCGAGCATGTGCGCCCTGACCGAGAG	1560
Db	121 CACATCGCCCGCAACTGCGCGGCCCGCGCAAGAGGGCTGCTGGAAGTGGGCAAGGAG	Qy	1555	GCCGAGCTGGAGTGGCGCGAGAAACCGCGAGATCTTGCAGCGCCCGTGTGACCGCGTGTAC	1614
Qy	181 GGCACACAGATGAAGAGTGAACGAGCGCGCCAGCCAACTTCTTCGCGAGAGACTGGCC	Db			
Db	181 GGCACACAGATGAAGAGTGAACGAGCGCGCCAGCCAACTTCTTCGCGAGAGACTGGCC	Qy			
Qy	241 TTCCCCAGAGGAGGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCCACCC	Db			
Db	241 TTCCCCAGAGGAGGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCCACCC	Qy			
Qy	301 AGCGCGAGCTGACAGGTGCGCGCGCAACCCCGCAGCGAGGCGCGCGCGCGAGCGCCAG	Db			
Db	301 AGCGCGAGCTGACAGGTGCGCGCGCAACCCCGCAGCGAGGCGCGCGCGCGAGCGCCAG	Qy			
Qy	361 GGCACCCGTGAATCTTCCCGCAGATCACCTGTGCGAGCGCCCGCTTGTGAGCATCAAGGTG	Db			
Db	361 GGCACCCGTGAATCTTCCCGCAGATCACCTGTGCGAGCGCCCGCTTGTGAGCATCAAGGTG	Qy			
Qy	421 GCGGCGCAGATCAAGGAGGCGCTGCTGACACCGCGCCCGACACCGTGTGGAGGAG	Db			
Db	421 GCGGCGCAGATCAAGGAGGCGCTGCTGACACCGCGCCCGACACCGTGTGGAGGAG				



```
Db 1561 GCGAGCTGGAGCTGCGCGAGAACCGCGAGATCTCTCGCGAGCCCGTGCACGGCGTGATC 1620
Qy 1615 TACGACCCCGACGAGACCTGTGTGCGGAGATCTCAAGAACGAGGCGCACGACGATGGAC 1674
Db 1621 TACGACCCCGACGAGACCTGTGTGCGGAGATCTCAAGAACGAGGCGCACGACGATGGAC 1680
Qy 1675 TACGAGATCTACCGAGGCGCTTCAAGAACCTGAGAGCGGCGACGATGCGCAAGATGCGC 1734
Db 1681 TACGAGATCTACCGAGGCGCTTCAAGAACCTGAGAGCGGCGACGATGCGCAAGATGCGC 1740
Qy 1735 ACCGCGCCACACGAGCTGAAGCAGCTGACCCGAGGCGGTCGACGAAGATGCGCATGGAG 1794
Db 1741 ACCGCGCCACACGAGCTGAAGCAGCTGACCCGAGGCGGTCGACGAAGATGCGCATGGAG 1800
Qy 1795 AGCATGCTGATCTGGGGCAAGACCCCAAGTTTCGCTGCCCATCCAGAGAGACCTGG 1854
Db 1801 AGCATGCTGATCTGGGGCAAGACCCCAAGTTTCGCTGCCCATCCAGAGAGACCTGG 1860
Qy 1855 GAGACCTGGTGGACCGACTACTGCGAGCGCCACCTGGATCCCGAGTGGAGTTCTGTGAC 1914
Db 1861 GAGACCTGGTGGACCGACTACTGCGAGCGCCACCTGGATCCCGAGTGGAGTTCTGTGAC 1920
Qy 1915 ACCCGCCCGCTGTTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGGCGCGAG 1974
Db 1921 ACCCGCCCGCTGTTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGGCGCGAG 1980
Qy 1975 ACCTTCTACGTGAGCGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2034
Db 1981 ACCTTCTACGTGAGCGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG 2040
Qy 2035 ACCGACCGGGCGCGCAGAGATCTGAGCTGACCGAGACCAACACGAGAGACCGAG 2094
Db 2041 ACCGACCGGGCGCGCAGAGATCTGAGCTGACCGAGACCAACACGAGAGACCGAG 2100
Qy 2095 CTGACGGGCATCCAGCTGGCCCTGCGAGACAGCGGCGAGGTTGAACATCGTGACCGAC 2154
Db 2101 CTGACGGGCATCCAGCTGGCCCTGCGAGACAGCGGCGAGGTTGAACATCGTGACCGAC 2160
Qy 2155 AGCCAGTACGGCTGGGATCATCATCGCGCCAGCCCGACGAGGAGGAGCTGTGTG 2214
Db 2161 AGCCAGTACGGCTGGGATCATCATCGCGCCAGCCCGACGAGGAGGAGCTGTGTG 2220
Qy 2215 AACGAGATCATGACGACCTGATCAAGAGGAGAGGTTGACCTGAGCTGGGTGCCGCC 2274
Db 2221 AACGAGATCATGACGACCTGATCAAGAGGAGAGGTTGACCTGAGCTGGGTGCCGCC 2280
Qy 2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334
Db 2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Qy 2335 GTGCTGTTCTTGGACGGATCATGATGGCGGATCTGATGATCTACAGTATGAGACGACCTG 2394
Db 2341 GTGCTGTTCTTGGACGGATCATGATGGCGGATCTGATGATCTACAGTATGAGACGACCTG 2400
Qy 2395 TACGTGGGACGGCGGCGCTAGGATCCATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2454
Db 2401 TACGTGGGACGGCGGCGCTAGGATCCATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2460
Qy 2455 TTC 2457
Db 2461 TTC 2463
```

RESULT 8  
ADCI3231  
ID ADCI3231 standard; DNA; 3930 BP.

XX ADCI3231;

AC ADCI3231;

XX 18-DEC-2003 (first entry)

DE DNA of HIV construct GagComplPolmutAtt\_C SEQ ID NO 10.

```
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX Human immunodeficiency virus.  
OS WO2003004620-A2.  
XX 16-JAN-2003.  
XX 05-JUL-2002; 2002WO-US021420.  
XX 05-JUL-2001; 2001US-0303192P.  
XX 31-AUG-2001; 2001US-0316860P.  
XX 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ;  
XX WPI; 2003-221593/21.  
XX New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.  
XX Disclosure; Fig 7; 301pp; English.  
XX The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intradermally, intranasally,  
CC subcutaneously, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.  
XX Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;  
SQ  
Query Match 99.1%; Score 2434.8; DB 10; Length 3930;  
Best Local Similarity 99.9%; Pred. No. 5.1e-293;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCGCGCAACATCTGATGAGCGCAGCAACTTCA 73  
Db 1487 TCGCGAGGCCATGAGCCAGGCCACCGCGCAACATCTGATGAGCGCAGCAACTTCA 1546  
Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGCCACATCGCCCGCA 133  
Db 1547 AGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGAGGGCCACATCGCCCGCA 1606  
Qy 134 ACTCGCGGCCCGCCGCAAGAGGGCTGCTGGAGTGGCGCAAGAGGGCCACCGAGTGA 193  
Db 1607 ACTCGCGGCCCGCCGCAAGAGGGCTGCTGGAGTGGCGCAAGAGGGCCACCGAGTGA 1666  
Qy 194 AGGACTGCACGAGCGCGCCAGGCCCAACTTCTTCCGAGGAGCTGGCTTCCCGCAGGGCA 253  
Db 1667 AGGACTGCACGAGCGCGCCAGGCCCAACTTCTTCCGAGGAGCTGGCTTCCCGCAGGGCA 1726  
Qy 254 AGGCCCCGAGTTCCTCCAGCGAGAGAACCCCGCGCAACAGCCCCCAACAGCGCGCGAGCTGC 313  
Db 1727 AGGCCCCGAGTTCCTCCAGCGAGAGAACCCCGCGCAACAGCCCCCAACAGCGCGCGAGCTGC 1786  
Qy 314 AGGTGCGGGCGACAAACCCCGCGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 1787 AGGTGCGGGCGACAAACCCCGCGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
Qy 374 TCCCCCAGATCACCTGTGTGGCAGCGGCCCTTGGTGAGCATCAAGGTGGCGCGCGCGCGCG 433
```

Db 1847 TCCCCAGATCACCTGTGGCAGGCCCCCTGGTGAGCATCAAGGTGGCGGCAGATCA 1906  
Qy 434 AGGAGGCCCTGCTGGACACCGGCGCCGACGACACCGTGTCTGGAGGAGATGAGCTTGCCCG 493  
Db 1907 AGGAGGCCCTGCTGGACTTCGGCGCCGACGACACCGTGTCTGGAGGAGATGAGCTTGCCCG 1966  
Qy 494 GCAAGTGGAAACCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGTGGCGCAGATCG 553  
Db 1967 GCAAGTGGAAACCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAGTGGCGCAGTACG 2026  
Qy 554 ACCAGATCCTCATCGAGATCTGGCGCAAGAGGCCATCGGCAACGCTGCTGATCGGCCCA 613  
Db 2027 ACCAGATCCTCATCGAGATCTGGCGCAAGAGGCCATCGGCAACGCTGCTGATCGGCCCA 2086  
Qy 614 CCCCCTGGAAATCATATCGGCGCAACATGCTGAACCCAGCTGGGTGCAACCTGGAATTC 673  
Db 2087 CCCCCTGGAAATCATATCGGCGCAACATGCTGAACCCAGCTGGGTGCAACCTGGAATTC 2146  
Qy 674 CCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGGAAGGCCCAAGG 733  
Db 2147 CCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGGAAGGCCCAAGG 2206  
Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793  
Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266  
Qy 794 TGAAGAGGAGGCGCAAGATCAACGAATCGGCGCCCGAGAACCCCTACAACACCCCTGT 853  
Db 2267 TGAAGAGGAGGCGCAAGATCAACGAATCGGCGCCCGAGAACCCCTACAACACCCCTGT 2326  
Qy 854 TCGCCATCAAGAAGAGGAGACGACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 913  
Db 2327 TCGCCATCAAGAAGAGGAGACGACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGA 2386  
Qy 914 ACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCAATCCCCCAACCGCGCGCTGA 973  
Db 2387 ACAGCGCACCCAGACTTCTGGAGGTGAGCTGGGCAATCCCCCAACCGCGCGCTGA 2446  
Qy 974 AGAAGAAGAGAGCGTGACCGTGTGGACGTGGCGAGCGCTACTTCAGCGTGCCCTGG 1033  
Db 2447 AGAAGAAGAGAGCGTGACCGTGTGGACGTGGCGAGCGCTACTTCAGCGTGCCCTGG 2506  
Qy 1034 ACAGGACTTCCGCAAGTACACCGCTTCAACATCCCCAGATCAACAAAGAGACCCCG 1093  
Db 2507 ACAGGACTTCCGCAAGTACACCGCTTCAACATCCCCAGATCAACAAAGAGACCCCG 2566  
Qy 1094 GCATCCGCTACAGTACAACTGTGCCCCAGGCTGGAGGGCAGCCCAAGCATCTTCC 1153  
Db 2567 GCATCCGCTACAGTACAACTGTGCCCCAGGCTGGAGGGCAGCCCAAGCATCTTCC 2626  
Qy 1154 AGACGAGCATGACCAAGATCTTGAGCGCTTTCGCGCCCGCAACCCCGAGATCGTGATCT 1213  
Db 2627 AGACGAGCATGACCAAGATCTTGAGCGCTTTCGCGCCCGCAACCCCGAGATCGTGATCT 2686  
Qy 1214 ACCAGGCCCCCTGTACGTGGGCGACGACCTGGAGATCGGCGAGCAGCCCGCCCAAGATCG 1273  
Db 2687 ACCAGGCCCCCTGTACGTGGGCGACGACCTGGAGATCGGCGAGCAGCCCGCCCAAGATCG 2746  
Qy 1274 AGGAGCTGGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGCAAGAGCAACGAGA 1333  
Db 2747 AGGAGCTGGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGCAAGAGCAACGAGA 2806  
Qy 1334 AGGAGCCCCCTTCTGCCCCATCGAGCTGCAACCCCGCAAGTGGACCGTGCAGGCCATCG 1393  
Db 2807 AGGAGCCCCCTTCTGCCCCATCGAGCTGCAACCCCGCAAGTGGACCGTGCAGGCCATCG 2866  
Qy 1394 AGCTGCCCGAGAGGAGCTGGACCGGTGAACGACATCCAGAACTGGTGGGCAAGCTGA 1453  
Db 2867 AGCTGCCCGAGAGGAGCTGGACCGGTGAACGACATCCAGAACTGGTGGGCAAGCTGA 2926  
Qy 1454 ACTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTGGCGAGCTGTGAAGCTGTGCGCG 1513

Db 2927 ACTGGGCGAGCGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAAGCTGCTGC 2986  
Qy 1514 GCGCAAGGCCCTTGACCGACATCGTGGCCCTTGACCGGAGGCGGAGCTGGAGCTGGCG 1573  
Db 2987 GCGCAAGGCCCTTGACCGACATCGTGGCCCTTGACCGGAGGCGGAGCTGGAGCTGGCG 3046  
Qy 1574 AGAACCGGAGATCTCTGGCGAGCCCTGCACGCGGTGTACTACGACCCCAAGCAAGACC 1633  
Db 3047 AGAACCGGAGATCTCTGGCGAGCCCTGCACGCGGTGTACTACGACCCCAAGCAAGACC 3106  
Qy 1634 TGGTGGCGGAGATCCAGAAGCAGGCGCAACGACGAGTGGACCTTACAGATCTTACAGGAGC 1693  
Db 3107 TGGTGGCGGAGATCCAGAAGCAGGCGCAACGACGAGTGGACCTTACAGATCTTACAGGAGC 3166  
Qy 1694 CTTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGCACGCCCCACACCAACGAGC 1753  
Db 3167 CTTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGCACGCCCCACACCAACGAGC 3226  
Qy 1754 TGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAAGAGCATCGTGAATCTGGGGCA 1813  
Db 3227 TGAAGCAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGGAAGAGCATCGTGAATCTGGGGCA 3286  
Qy 1814 AGACCCCAAGTTCCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 1873  
Db 3287 AGACCCCAAGTTCCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACT 3346  
Qy 1874 ACTGGCAGGGCCACTGTGATCCCGAGTGGGAGTTCTGTAAACACCCCGCCCTGCTGAGAGC 1933  
Db 3347 ACTGGCAGGGCCACTGTGATCCCGAGTGGGAGTTCTGTAAACACCCCGCCCTGCTGAGAGC 3406  
Qy 1934 TGTGGTACCAAGCTGGAGAAAGAGCCCATCATCGGCGCCGAGACCTTCTAAGTGAACGCGC 1993  
Db 3407 TGTGGTACCAAGCTGGAGAAAGAGCCCATCATCGGCGCCGAGACCTTCTAAGTGAACGCGC 3466  
Qy 1994 CCGCAACCCGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGCGGCGGCGGCGAG 2053  
Db 3467 CCGCAACCCGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGCGGCGGCGGCGAG 3526  
Qy 2054 AGATCGTGAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG 2113  
Db 3527 AGATCGTGAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG 3586  
Qy 2114 CCTTGAGGACAGCGCGCAGCGAGGTGAACATCGTGACCGACAGCAAGTACGCCCTGGGCA 2173  
Db 3587 CCTTGAGGACAGCGCGCAGCGAGGTGAACATCGTGACCGACAGCAAGTACGCCCTGGGCA 3646  
Qy 2174 TCATCCAGGCCCGACCGCAAGAGCGAGCGAGCTGTAACCAAGATCATCGAGCAGC 2233  
Db 3647 TCATCCAGGCCCGACCGCAAGAGCGAGCGAGCTGTAACCAAGATCATCGAGCAGC 3706  
Qy 2234 TGATCAAGAAAGGAGAAAGGTGTACTGTAGCTGGGTGCCCGCCCAAGGGCATCGGCGGCA 2293  
Db 3707 TGATCAAGAAAGGAGAAAGGTGTACTGTAGCTGGGTGCCCGCCCAAGGGCATCGGCGGCA 3766  
Qy 2294 ACAGACGATCGCAAGCTGGTGAAGGAGGCAATCCGCAAGGTGCTGTCTTGAGACGCGCA 2353  
Db 3767 ACAGACGATCGCAAGCTGGTGAAGGAGGCAATCCGCAAGGTGCTGTCTTGAGACGCGCA 3826  
Qy 2354 TCGATGGCGGCATCGTGAATCTACAGTACATGGAACGACCTGTAGCTGGGCGAGCGCGGCC 2413  
Db 3827 TCGATGGCGGCATCGTGAATCTACAGTACATGGAACGACCTGTAGCTGGGCGAGCGCGGCC 3886  
Qy 2414 CTAGGATCGAATTAAGCTTCCCGGGCTAGACCGGT 2451  
Db 3887 CTAGGATCGAATTAAGCTTCCCGGGCTAGACCGGT 3924

RESULT 9

ADCL3232

ID ADCL3232 standard; DNA; 3930 BP.

XX AC

XX ADCL3232;

XX AC



QY	1454	ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAGAGCTGCTGC	1513
DB	2927	ACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAGAGCTGCTGC	2986
QY	1514	GCGCCAAGGCCCTTACCGGACATCTGTGCCCTTGACCGAGGAGGCCGAGCTGGAGCTGG	1573
DB	2987	GCGCCAAGGCCCTTACCGGACATCTGTGCCCTTGACCGAGGAGGCCGAGCTGGAGCTGG	3046
QY	1574	AGAACCGCGAGATCCTGGCGAGCCCGTGACCGCGGTGTACTACGCCCCAGCAGAGACC	1633
DB	3047	AGAACCGCGAGATCCTGGCGAGCCCGTGACCGCGGTGTACTACGCCCCAGCAGAGACC	3106
QY	1634	TGTTGGCCGAGATCCAGAAGCAGAGGCCACGACCACTGAGTGGACCTTACCAGATCTTACC	1693
DB	3107	TGTTGGCCGAGATCCAGAAGCAGAGGCCACGACCACTGAGTGGACCTTACCAGATCTTACC	3166
QY	1594	CTTTCAAGAACCTTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCCCAACCAACGAG	1753
DB	3167	CTTTCAAGAACCTTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCCCAACCAACGAG	3226
QY	1754	TGAAGCAGCTGACCGAGGCCGTGAGAAGATTCGCCCATTGGAGAGCATCGTGATCTTGGG	1813
DB	3227	TGAAGCAGCTGACCGAGGCCGTGAGAAGATTCGCCCATTGGAGAGCATCGTGATCTTGGG	3286
QY	1814	AGACCCCAAGTTCCGCTGCCATTCAGAAGGAGACCTGGGAGACCTTGGTGGACCGACT	1873
DB	3287	AGACCCCAAGTTCCGCTGCCATTCAGAAGGAGACCTGGGAGACCTTGGTGGACCGACT	3346
QY	1874	ACTGGCAGGCCACTGTGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCCCTGGTGAAG	1933
DB	3347	ACTGGCAGGCCACTGTGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCCCTGGTGAAG	3406
QY	1934	TGTGGTACCAGCTGGGAAGGAGCCATCATTCGGCGCCGAGACCTTCTACGTGGACGGCG	1993
DB	3407	TGTGGTACCAGCTGGGAAGGAGCCATCATTCGGCGCCGAGACCTTCTACGTGGACGGCG	3466
QY	1994	CCGCCAACCGGAGACCAAGATTCGGCAAGGCGCGGTACGTGAACGACCGGGGCCGCGAGA	2053
DB	3467	CCGCCAACCGGAGACCAAGATTCGGCAAGGCGCGGTACGTGAACGACCGGGGCCGCGAGA	3526
QY	2054	AGATCGTGAGCTTGACCGAGACCAACACAGAAAGACCGAGCTGACGGCCCATCCAGCTGG	2113
DB	3527	AGATCGTGAGCTTGACCGAGACCAACACAGAAAGACCGAGCTGACGGCCCATCCAGCTGG	3586
QY	2114	CCCTGCAGGACAGCGGCAGCGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCA	2173
DB	3587	CCCTGCAGGACAGCGGCAGCGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCA	3646
QY	2174	TCACTCAGGCCACGCGGACAGAGCGAGCTGTGTGAACCGAGATCATTCGAGCAGC	2233
DB	3647	TCACTCAGGCCACGCGGACAGAGCGAGCTGTGTGAACCGAGATCATTCGAGCAGC	3706
QY	2234	TGATCAAGAAGGAGAAGGTGTACTGTGAGCTGGGTGCCCGGCCCAAGAGGATCGGGCGGCA	2293
DB	3707	TGATCAAGAAGGAGAAGGTGTACTGTGAGCTGGGTGCCCGGCCCAAGAGGATCGGGCGGCA	3766
QY	2294	ACGAGCAGATTCGACAGCTGGTGAAGAGGCGATCCGCAAGAGGTGCTGTTCTCTGAACGGCA	2353
DB	3767	ACGAGCAGATTCGACAGCTGGTGAAGAGGCGATCCGCAAGAGGTGCTGTTCTCTGAACGGCA	3826
QY	2354	TCGATGGCGGCATCGTGATCTTACCAAGTACATGGAACGACCTGTACGTGGGACGCGCGGCC	2413
DB	3827	TCGATGGCGGCATCGTGATCTTACCAAGTACATGGAACGACCTGTACGTGGGACGCGCGGCC	3886
QY	2414	CTAGGATCGATATAAAGCTTCCCGGGGCTTAGCACCGGT	2451
DB	3887	CTAGGATCGATATAAAGCTTCCCGGGGCTTAGCACCGGT	3924

RESULT 10  
ACA03591  
ID ACA0

XX	ACA03591;	
AC		
XX		
XX	22-MAY-2003 (first entry)	
DT		
XX		
XX	Synthetic DNA encoding immunogenic HIV peptide #74.	
DE		
XX		
KW	Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;	
KW	gene therapy; packaging cell line; humoral immune response;	
KW	cellular immune response; gene delivery vector; DNA immunisation; ds.	
XX		
OS	Synthetic.	
XX		
FN	WO2003004657-A1.	
XX		
XX	16-JAN-2003.	
PD		
XX		
XX	05-JUL-2002; 2002WO-US021421.	
PF		
XX		
XX	05-JUL-2001; 2001US-0303192P.	
PR		
PR	31-AUG-2001; 2001US-0316860P.	
PR	16-JAN-2002; 2002US-0349728P.	
PR	16-JAN-2002; 2002US-0349733P.	
PR	16-JAN-2002; 2002US-0349871P.	
PR		
XX	(CHIR ) CHIRON CORP.	
PA		
XX		
XX		
PI	Zur Megede J, Barnett SW, Lian Y;	
XX		
XX	WPI; 2003-221602/21.	
DR		
XX		
PT	New synthetic polynucleotides encoding antigenic HIV type B and/or type C	
PT	polypeptides, useful as immunogenic compositions or vaccines for	
PT	generating humoral or cellular immune responses against HIV in a subject,	
PT	especially humans.	
PS		
XX	Example 1; Fig 79; 262pp; English.	
XX		
CC	The invention describes a synthetic polynucleotide encoding 2 or more	
CC	immunogenic HIV polypeptides, where at least 2 of the polypeptides are	
CC	derived from different HIV subtypes. The polynucleotide is useful for	
CC	immunisation, generation of packaging cell lines, or production of HIV	
CC	polypeptides. The polynucleotide and its encoded proteins are useful as	
CC	immunogenic compositions or vaccines for generating humoral or cellular	
CC	immune responses against HIV in a subject, or for inducing neutralising	
CC	antibodies against HIV. The gene delivery vector comprising the	
CC	polynucleotide is also useful for DNA immunisation of, or for generating	
CC	an immune response (e.g. a humoral or cellular immune response) in, a	
CC	subject such as a mammal, particularly a human. This sequence encodes a	
CC	human immunodeficiency virus immunogenic peptide	
XX		
XX	Sequence 5184 BP: 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;	

	Query Match	99.1%	Score	2434.8	DB 8	Length	5184	
	Best Local Similarity	99.9%	Pred. No.	4.9e-293				
	Matches	2436	Conservative	0	Mismatches	2	Indels	0
								Gaps
								0
Qy	14	TGCGCAGGCGCATGAGCAGGCCACACAGCGCCCAACATCTCTGATGCGAGCGAGCAACTTCA	73					
Db	2741	TGCGCGAGGCATGAGCCAGGCCACACAGCGCCCAACATCTCTGATGCGAGCGAGCAACTTCA	2800					
Qy	74	AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTTGCGGCAAGGAGGGGCCACATCGCCCCGCA	133					
Db	2801	AGGGCCCCAAGCGCATCATCAAGTGCTTCAACTTGCGGCAAGGAGGGGCCACATCGCCCCGCA	2860					
Qy	134	ACTGCGCGCGCCCCCGCCAGGAAGGGCTGCTGGGAAGTGCGGCAAGGAGGGGCCACACAGATGA	193					
Db	2861	ACTGCGCGCGCCCCCGCCAGGAAGGGCTGCTGGGAAGTGCGGCAAGGAGGGGCCACACAGATGA	2920					
Qy	194	AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGAGGACCTTGGCGCTTCCCCCAGGGCA	253					
Db	2921	AGGACTGCACCGAGCGCAGGCCAACTTCTTCGCGAGGACCTTGGCGCTTCCCCCAGGGCA	2980					



Db 5141 CTAGATCGATTAAAGCTTCCGGGGCTAGCACGGT 5178

RESULT 11

ADCI3279

ID ADCI3279 standard; DNA; 5184 BP.

XX

AC ADCI3279;

XX

DT 18-DEC-2003 (first entry)

XX

DE DNA of HIV construct TatRevWefagCpolIna\_C SEQ ID NO 58.

XX

KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

XX

KW Rev; HIV type C; Gene therapy; vaccine; immunisation; HIV; ds.

XX

OS Human immunodeficiency virus.

XX

PN WO2003004620-A2.

XX

PD 16-JAN-2003.

XX

PF 05-JUL-2002; 2002WO-US021420.

XX

PR 05-JUL-2001; 2001US-0303192P.

PR

PR 31-AUG-2001; 2001US-0316860P.

PR

PR 16-JAN-2002; 2002US-0349871P.

XX

XX (CHIR ) CHIRON CORP.

PA (UYST-) UNIV STELLENBOSCH.

XX

PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX

XX WPI; 2003-221593/21.

XX

XX New expression cassette comprising a polynucleotide sequence encoding a

PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

PT Prot, or Rev polypeptide, useful for immunization, or generating

PT packaging cell lines.

XX

XX Disclosure; Fig 55; 301pp; English.

XX

CC The invention relates to a novel expression cassette comprising a

CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,

CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

CC expression cassette can be used to treat HIV type C by gene therapy or

CC used in the development of a vaccine. The gene delivery vector is

CC administered intramuscularly, intracutaneously, intravenously,

CC subcutaneously, intradermally, transdermally, intravaginally,

CC intrarectally, orally or intravenously. The expression cassette is useful

CC for immunisation, generating packaging cell lines and producing HIV

CC polypeptides. This polynucleotide sequence represents the DNA of an HIV

CC Type C related sequence of the invention.

XX

SQ Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match

Best Local Similarity 99.1%; Score 2434.8; DB 10; Length 5184;

Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCGCATGAGCGAGGCCACACGAGCGCAATCTTGATGAGCGAGCAACTTCA 73

Db

Qy 74 AGGCCCCCAGCGCATCATCAAGTGTCTCACTGGCGAGGAGGCGGCATCGCCCGCA 133.

Db

Db 2801 AGGCCCCCAGCGCATCATCAAGTGTCTCACTGGCGAGGAGGCGGCATCGCCCGCA 2860

Qy 134 ACTCGCGCGCCCCCGGAGAGGGCTGTGGAAGTGGCGAGGAGGCGGCACAGATGA 193

Db

Db 2861 ACTCGCGCGCCCCCGGAGAGGGCTGTGGAAGTGGCGAGGAGGCGGCACAGATGA 2920

Qy 194 AGGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGAGCTTCCCGCTTCCCGAGGGCA 253

Db 2921 AGGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGAGCTTCCCGCTTCCCGAGGGCA 2980

Qy

254 AGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCAACAGCCCGCGAGCTGC 313

Db

2981 AGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCAACAGCCCGCGAGCTGC 3040

Qy

314 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373

Db

3041 AGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100

Qy

374 TCCCGCAGATCACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433

Db

3101 TCCCGCAGATCACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160

Qy

434 AGGAGGCGCTGTGTGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493

Db

3161 AGGAGGCGCTGTGTGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3220

Qy

494 GCAAGTGGAAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553

Db

3221 GCAAGTGGAAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3280

Qy

554 ACCAGATCCTGATCGAGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613

Db

3281 ACCAGATCCTGATCGAGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3340

Qy

614 CCCCCTGTGAACATCATCG 673

Db

3341 CCCCCTGTGAACATCATCG 3400

Qy

674 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCGCGCGCGCGCGCGCGCGCGCG 733

Db

3401 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCGCGCGCGCGCGCGCGCGCGCG 3460

Qy

734 TGAAGCAGTGGCG 793

Db

3461 TGAAGCAGTGGCG 3520

Qy

794 TGGAGAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853

Db

3521 TGGAGAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3580

Qy

854 TCGCCATCAAGAGAAGGAGCAGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 913

Db

3581 TCGCCATCAAGAGAAGGAGCAGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3640

Qy

914 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCG 973

Db

3641 ACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCG 3700

Qy

974 AGAAGAGAGAGCGTGACCGTGTCTGGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1033

Db

3701 AGAAGAGAGAGCGTGACCGTGTCTGGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3760

Qy

1034 ACAGAGACTTCCGCAAGTACACCGCGCTTCCACATCCCGCGCGCGCGCGCGCGCGCGCGCG 1093

Db

3761 ACAGAGACTTCCGCAAGTACACCGCGCTTCCACATCCCGCGCGCGCGCGCGCGCGCGCGCG 3820

Qy

1094 GCATCCGCTACCAAGTACACCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1153

Db

3821 GCATCCGCTACCAAGTACACCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3880

Qy

1154 AGAGCAGCATGACCAAGATCCTTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1213

Db

3881 AGAGCAGCATGACCAAGATCCTTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3940

Qy

1214 ACCAGGCGCGCGCTGTGTGAGCG 1273

Db

3941 ACCAGGCGCGCGCTGTGTGAGCG 4000

Qy

1274 AGGAGTGGCGCAAGCAGCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1333

Db 4001 AGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCAACACCCCGCACAAAGACACCAGA 4060  
Qy 1334 AGGAGCCCCCTTCTTCTGCGCATCGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCG 1393  
Db 4061 AGGAGCCCCCTTCTTCTGCGCATCGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCG 4120  
Qy 1394 AGCTGCCCGAGAGAGAGCTGGACCGTGAAGCAGATCCAGAACTGGTGGGCAAGCTGA 1453  
Db 4121 AGCTGCCCGAGAGAGAGCTGGACCGTGAAGCAGATCCAGAACTGGTGGGCAAGCTGA 4180  
Qy 1454 ACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCG 1513  
Db 4181 ACTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCG 4240  
Qy 1514 GGGCCAAAGCCCTGACCGACATCGTGCCTTACCGAGAGGCCGAGCTGGAGCTGGCGG 1573  
Db 4241 GGGCCAAAGCCCTGACCGACATCGTGCCTTACCGAGAGGCCGAGCTGGAGCTGGCGG 4300  
Qy 1574 AGAACCGGAGATCTTGGCGGAGCCGCTGACGGGTGTACTACGACCCCAAGCAAGGACC 1633  
Db 4301 AGAACCGGAGATCTTGGCGGAGCCGCTGACGGGTGTACTACGACCCCAAGCAAGGACC 4360  
Qy 1634 TGGTGGCCGAGATCCAGAGCAGGCGCACGACCTAGTGGACCTTACAGATCTTACAGGAGC 1693  
Db 4361 TGGTGGCCGAGATCCAGAGCAGGCGCACGACCTAGTGGACCTTACAGATCTTACAGGAGC 4420  
Qy 1694 CTTTCAAGAACCTGAAGACCGCAGTAGTACGCCAAGATGGGCACCGCCACACCAACGAGG 1753  
Db 4421 CTTTCAAGAACCTGAAGACCGCAGTAGTACGCCAAGATGGGCACCGCCACACCAACGAGC 4480  
Qy 1754 TGAAGCAGCTGACCGAGGCGGTGCAGAGATGCGCATGGAGAGCATCGTGTCTGGGCA 1813  
Db 4481 TGAAGCAGCTGACCGAGGCGGTGCAGAGATGCGCATGGAGAGCATCGTGTCTGGGCA 4540  
Qy 1814 AGACCCCAAGTTCCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 1873  
Db 4541 AGACCCCAAGTTCCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 4600  
Qy 1874 ACTGGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGGTGAAGC 1933  
Db 4601 ACTGGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGGTGAAGC 4660  
Qy 1934 TGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGG 1993  
Db 4661 TGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGG 4720  
Qy 1994 CGGCCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGGGCAGA 2053  
Db 4721 CGGCCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGGGCAGA 4780  
Qy 2054 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGTGCAGGGCCATCCAGCTGG 2113  
Db 4781 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGTGCAGGGCCATCCAGCTGG 4840  
Qy 2114 CCTTCAGAGACCGGCGAGGAGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCA 2173  
Db 4841 CCTTCAGAGACCGGCGAGGAGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCA 4900  
Qy 2174 TCATCCAGGCCCGCCGCAAGAGCGAGGAGCTGGTGAACCAAGATCATCGAGCAGC 2233  
Db 4901 TCATCCAGGCCCGCCGCAAGAGCGAGGAGCTGGTGAACCAAGATCATCGAGCAGC 4960  
Qy 2234 TCATCAAGAGAGAGGTGTACCTGAGCTGGGTGGTCCGCCCAAGGGCATCCGGGCA 2293  
Db 4961 TCATCAAGAGAGAGGTGTACCTGAGCTGGGTGGTCCGCCCAAGGGCATCCGGGCA 5020  
Qy 2294 AGGAGCAGATCGACAAGCTGGTGAAGGGGATCCGCAAGGTGTCTTCTGGAGCGCA 2353  
Db 5021 AGGAGCAGATCGACAAGCTGGTGAAGGGGATCCGCAAGGTGTCTTCTGGAGCGCA 5080  
Qy 2354 TCGATGGGCGCATCGTGTACTACAGTACATCGACGACCTGTACTGGGCGAGCGCGCC 2413  
Db 5081 TCGATGGGCGCATCGTGTACTACAGTACATCGACGACCTGTACTGGGCGAGCGCGCC 5140

Qy 2414 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 5141 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 5178

RESULT 12  
ACA03547  
ID ACA03547 standard; DNA; 2457 BP.  
XX ACA03547;  
XX AC  
XX DT 22-MAY-2003 (first entry)  
XX XX  
DE Synthetic DNA ecoding immunogenic HIV peptide #30.  
XX Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
KW gene therapy; packaging cell line; humoral immune response;  
KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
XX Synthetic.  
XX OS  
XX PN WO2003004657-A1.  
XX PD 16-JAN-2003.  
XX PF 05-JUL-2002; 2002WO-US021421.  
XX PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349728P.  
PR 16-JAN-2002; 2002US-0349793P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX (CHIR ) CHIRON CORP.  
XX PA  
XX PI Zur Megede J, Barnett SW, Lian Y;  
XX WPI; 2003-221602/21.  
XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
PT polypeptides, useful as immunogenic compositions or vaccines for  
PT generating humoral or cellular immune responses against HIV in a subject,  
PT especially humans.  
XX Example 1; Fig 35; 262pp; English.  
XX The invention describes a synthetic polynucleotide encoding 2 or more  
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
CC derived from different HIV subtypes. The polynucleotide is useful for  
CC immunisation, generation of packaging cell lines, or production of HIV  
CC polypeptides. The polynucleotide and its encoded proteins are useful as  
CC immunogenic compositions or vaccines for generating humoral or cellular  
CC immune responses against HIV in a subject, or for inducing neutralising  
CC antibodies against HIV. The gene delivery vector comprising the  
CC polynucleotide is also useful for DNA immunisation of, or for generating  
CC an immune response (e.g. a humoral or cellular immune response) in, a  
CC subject such as a mammal, particularly a human. This sequence encodes a  
CC human immunodeficiency virus immunogenic peptide  
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.8%; Score 2428.6; DB 8; Length 2457;  
Best Local Similarity 99.6%; Fred. No. 3.2e-292;  
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 GTCGACGCCACCATGCGCCGAGCCCATGAGCCAGGCCACCGCCCAACATCTCTGATGCAG 60  
Db 1 GTCGACGCCACCATGCGCCGAGCCCATGAGCCAGGCCACCGCCCAACATCTCTGATGCAG 60

Qy 61 CGCAGCAACTTCAAGGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120  
Db 61 CGCAGCAACTTCAAGGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120



Qy 121 CACATCGCCCGCAACTGCGCGCCGCCCGCAAGAGGGCTGCTGAAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCCGCAACTGCGCGCGCCGCCCGCAAGAGGGCTGCTGAAAGTGGCGCAAGGAG 180  
Qy 181 GGCACACAGATGAAAGGACTGACAGAGCGCCAGAGCAACTTCTTCCGCGAGGACTGGCC 240  
Db 181 GGCACACAGATGAAAGGACTGACAGAGCGCCAGAGCAACTTCTTCCGCGAGGACTGGCC 240  
Qy 241 TTCCCGCAGGCAAGGCGCGGAGTTCCCGCAGCAGAGCAAGACCGCGCCCAACAGAGCCCAACC 300  
Db 241 TTCCCGCAGGCAAGGCGCGGAGTTCCCGCAGCAGAGCAAGACCGCGCCCAACAGAGCCCAACC 300  
Qy 301 AGCGCGAGCTGCAAGGTGGCGGAGCAAAACCCCGCAGCAGAGCGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTGCAAGGTGGCGGAGCAAAACCCCGCAGCAGAGCGCGCGCGAGCGCGCCAG 360  
Qy 361 GGCACCTGAACTTCCCCAGATCAACCTGTGGCAGCGCCCGCTGGTAGCATCAAGTG 420  
Db 361 GGCACCTGAACTTCCCCAGATCAACCTGTGGCAGCGCCCGCTGGTAGCATCAAGTG 420  
Qy 421 GCGCGCAGATCAAGGAGCGCTGCTGACACCGCGCGCAGACACCGTCTGAGGAG 480  
Db 421 GCGCGCAGATCAAGGAGCGCTGCTGACACCGCGCGCAGACACCGTCTGAGGAG 480  
Qy 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAG 540  
Db 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAG 540  
Qy 541 GTGCGCGATGACGACAGATCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCAACCGTG 600  
Db 541 GTGCGCGATGACGACAGATCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCAACCGTG 600  
Qy 601 CTGATCGGCCCGCCCGCTGAAATCATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCCCGCCCGCTGAAATCATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Qy 661 ACCCTGAACTTCCCCCATCAGCGCCCATCGAGACCGTGCCTGAAAGTGAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCCCATCAGCGCCCATCGAGACCGTGCCTGAAAGTGAAGCCCGGCATG 720  
Qy 721 GACGCGCCAAAGTGAAGCAGTGGCGCTGACCGAGAGGAGATCAAGGCGCTGACCGCC 780  
Db 721 GACGCGCCAAAGTGAAGCAGTGGCGCTGACCGAGAGGAGATCAAGGCGCTGACCGCC 780  
Qy 781 ATCTGCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAGAACCCCTAC 840  
Qy 841 AACACCCCGCTGTTCCGCATCAAGAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGCTGTTCCGCATCAAGAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCCGCGAGCTGAACAAGCGACCCAGACTTCTGGAGAGTGCAGTGGGCAATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAGCGACCCAGACTTCTGGAGAGTGCAGTGGGCAATCCCCAC 960  
Qy 961 CCGCGCGCTGAAGAAGAGAGGCTGACCGTGTGGACGTGGGCGAGCGCTACTTC 1020  
Db 961 CCGCGCGCTGAAGAAGAGAGGCTGACCGTGTGGACGTGGGCGAGCGCTACTTC 1020  
Qy 1021 AGCGTGGCCCTGAGCAGGAGCTTCCGCAAGTACACCGCTTTCACATCCCCAGCATCAAC 1080  
Db 1021 AGCGTGGCCCTGAGCAGGAGCTTCCGCAAGTACACCGCTTTCACATCCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGATCCGCTACAGTACAACTGTGCCCCAGAGGCTGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGCGATCCGCTACAGTACAACTGTGCCCCAGAGGCTGAAGGGCAGC 1140  
Qy 1141 CCCAGCATCTTCCAGAGCAGATGACCAAGATCTGGAGCCCTTCCGCGCGCGCAACCCC 1200  
Db 1141 CCCAGCATCTTCCAGAGCAGATGACCAAGATCTGGAGCCCTTCCGCGCGCGCAACCCC 1200

Qy 1201 GAGATCGTGATCTTACCAAGGCGCCCGCTGTATCGTGGCGAGCGACTGGAGATCGGCGCAGCAC 1260  
Db 1201 GAGATCGTGATCTTACCAAGGCGCCCGCTGTATCGTGGCGAGCGACTGGAGATCGGCGCAGCAC 1260  
Qy 1261 CGCGCCAGATCGAGGAGCTGCGCAAGCAGCTGCGCTGGGCTTCCACACCCCGCAC 1320  
Db 1261 CGCGCCAGATCGAGGAGCTGCGCAAGCAGCTGCGCTGGGCTTCCACACCCCGCAC 1320  
Qy 1321 AAGAAGCACCAAGAAGGAGCGCCCGCTTCTGCGCCAT-----CGAGCTGCACCCCGACAAG 1374  
Db 1321 AAGAAGCACCAAGAAGGAGCGCCCGCTTCTGCTGGATGGCTACGAGCTGCACCCCGACAAG 1380  
Qy 1375 TGGACCGTGCAGCCCATCGAGCTGCCGAGAAAGAGAGCTGGAACGCTGAAACGACATCCAG 1434  
Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAAAGAGAGCTGGAACGCTGAAACGACATCCAG 1440  
Qy 1435 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGCATCAAGGTGCGCCAG 1494  
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGCATCAAGGTGCGCCAG 1500  
Qy 1495 CTGTGCAAGTGTGTCGCGCGCCCAAGGCCCTGACCGACATCTGTGCCCTGACCGAGGAG 1554  
Db 1501 CTGTGCAAGTGTGTCGCGCGCCCAAGGCCCTGACCGACATCTGTGCCCTGACCGAGGAG 1560  
Qy 1555 GCCGAGCTGAGCTGGCGCGAGAACCCGCGAGATCTGCGCGAGCCCGTGCAACGCGCTGTAC 1614  
Db 1561 GCCGAGCTGAGCTGGCGCGAGAACCCGCGAGATCTGCGCGAGCCCGTGCAACGCGCTGTAC 1620  
Qy 1615 TAGCACCCGAGCAAGGACTGTGGCGCGAGATCTCAGAAAGCAGGCGCCAGCAGTGGAC 1674  
Db 1621 TAGCACCCGAGCAAGGACTGTGGCGCGAGATCTCAGAAAGCAGGCGCCAGCAGTGGAC 1680  
Qy 1675 TACCAGATCTTACCAAGGAGCCCTTCAAGAACTGGAAGACCGCAAGTACGCGCAAGTGGC 1734  
Db 1681 TACCAGATCTTACCAAGGAGCCCTTCAAGAACTGGAAGACCGCGCAAGTACGCGCAAGTGGC 1740  
Qy 1735 ACCGCGCCACCAACGACGTAAGCAGCTGACCGAGGCGCGTGCAAGAGATCGCCATGGAG 1794  
Db 1741 ACCGCGCCACCAACGACGTAAGCAGCTGACCGAGGCGCGTGCAAGAGATCGCCATGGAG 1800  
Qy 1795 AGCATCTGTATCTGGGCGCAAGACCCCGCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1854  
Db 1801 AGCATCTGTATCTGGGCGCAAGACCCCGCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1860  
Qy 1855 GAGACCTGTGTGACCCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTCCTGAAC 1914  
Db 1861 GAGACCTGTGTGACCCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTCCTGAAC 1920  
Qy 1915 ACCCGCGCTGTGTGAAGCTGTGGTACCGAGTGGAGAGGAGCCCATCATCGGCGCGAG 1974  
Db 1921 ACCCGCGCTGTGTGAAGCTGTGGTACCGAGTGGAGAGGAGCCCATCATCGGCGCGAG 1980  
Qy 1975 ACCTTCTAGTGAAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2034  
Db 1981 ACCTTCTAGTGAAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040  
Qy 2035 ACCGAGCGCGCGCGCGAGAGATCGTGACCTGACCGAGACCAACCAAGAGAGAGCGAG 2094  
Db 2041 ACCGAGCGCGCGCGCGAGAGATCGTGACCTGACCGAGACCAACCAAGAGAGAGCGAG 2100  
Qy 2095 CTGAGGCGCATCCAGCTGGCGCCCTGCGAGACAGCGCGAGCGAGTGAACATCGTACCGAC 2154  
Db 2101 CTGAGGCGCATCCAGCTGGCGCCCTGCGAGACAGCGCGAGCGAGTGAACATCGTACCGAC 2160  
Qy 2155 AGCCAGTACGCGCTTGGGCGATCATCCAGGCCAGCCCGCAAGAGCGAGGAGCGTGGTG 2214  
Db 2161 AGCCAGTACGCGCTTGGGCGATCATCCAGGCCAGCCCGCAAGAGCGAGGAGCGTGGTG 2220  
Qy 2215 AACAGATCATCCAGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGTCCCGGCC 2274  
Db 2221 AACAGATCATCCAGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGTCCCGGCC 2280  
Qy 2275 CACAAGGGCATCGCGCGCAACGAGCAGATCGAACAGCTGGTGGAGCAAGGGGCATCCGCAAG 2334

Db 2281 CACAAGGGCATCGGGCAACAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340  
QY 2335 GTGCTGTTCTTGAGACGGCATCGATGGCGGCATCGTGATCTACAGTACATGAGACACCTG 2394  
Db 2341 GTGCTGTTCTTGAGACGGCATCGATGGCGGCATCGTGATCTACAGTACATGAGACACCTG 2400  
QY 2395 TACGTGGGAGCGGGCGGCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2451  
Db 2401 TACGTGGGAGCGGGCGGCTAGGATCGATTAAAGCTTCCCGGGGTAGCACCGGT 2457

RESULT 13  
ADCL3265  
ID ADCL3265 standard; DNA; 2457 BP.  
XX  
AC ADC13265;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA of HIV construct p2Pol-opt-YM\_C SEQ ID NO 44.  
XX  
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
XX  
OS Human immunodeficiency virus.  
XX  
PN WO2003004620-A2.  
XX  
PD 16-JAN-2003.  
XX  
PF 05-JUL-2002; 2002WO-US021420.  
XX  
PR 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (UYST-) UNIV STELLENBOSCH.  
XX  
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
XX WPI; 2003-221593/21.  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating  
PT packaging cell lines.

XX Disclosure; Fig 41; 301pp; English.  
XX  
CC The invention relates to a novel expression cassette comprising a  
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
CC expression cassette can be used to treat HIV type C by gene therapy or  
CC used in the development of a vaccine. The gene delivery vector is  
CC administered intramuscularly, intramusosally, intranasally,  
CC subcutaneously, intradermally, transdermally, intravaginally,  
CC intrarectally, orally or intravenously. The expression cassette is useful  
CC for immunisation, generating packaging cell lines and producing HIV  
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
CC Type C related sequence of the invention.

XX Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;  
SQ  
Query Match 98.8%; Score 2428.6; DB 10; Length 2457;  
Best Local Similarity 99.6%; Pred. No. 3.2e-292;  
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 1 GTGAGCGCCACCATGGCGGCGCATAGCGCCAGGCGCACAGCGCCAAACATCTGTATGAG 60  
Db 1 GTGAGCGCCACCATGGCGGCGCATAGCGCCAGGCGCACAGCGCCAAACATCTGTATGAG 60

QY 61 CGCAGCAATTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGTGGGCAAGAGGGC 120  
Db 61 CGCAGCAATTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGTGGGCAAGAGGGC 120  
QY 121 CACATCGCCCCCAACTGCTCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGGGCAAGGAG 180  
Db 121 CACATCGCCCCCAACTGCTCCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGGGCAAGGAG 180  
QY 181 GGCACACAGATGAGGAGTGCACAGAGCGCCAGGCAATCTTCTTCGCGAGGACCTGGCC 240  
Db 181 GGCACACAGATGAGGAGTGCACAGAGCGCCAGGCAATCTTCTTCGCGAGGACCTGGCC 240  
QY 241 TTCCCCAGGCGCAAGCGCCGCGAGTTTCCCGAGCAGCAGAAACCGCGCCAAACAGACCCACC 300  
Db 241 TTCCCCAGGCGCAAGCGCCGCGAGTTTCCCGAGCAGCAGAAACCGCGCCAAACAGACCCACC 300  
QY 301 AGCGCGAGCTGCAGAGTGCAGGGGACAAACCCCGCAGCGAGCGCGCGCCAGGCCCAG 360  
Db 301 AGCGCGAGCTGCAGAGTGCAGGGGACAAACCCCGCAGCGAGCGCGCGCGCCAGGCCCAG 360  
QY 361 GGCACCTTGAATTTCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGTG 420  
Db 361 GGCACCTTGAATTTCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGTG 420  
QY 421 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGCGCGCAGCACACCGTGTGTGAGGAG 480  
Db 421 GCGGCCAGATCAAGGAGGCCCTGCTGGACACCGCGCGCAGCACACCGTGTGTGAGGAG 480  
QY 481 ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGGCTTCACTAAG 540  
Db 481 ATGAGCTTCCCGGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGGCTTCACTAAG 540  
QY 541 GTGCGCCAGTACAGCAGATCTGTAGATCTGCGGCAAGAGGCCCATCGGCAACGCTG 600  
Db 541 GTGCGCCAGTACAGCAGATCTGTAGATCTGCGGCAAGAGGCCCATCGGCAACGCTG 600  
QY 601 CTGATCGGGCCCCACCGCGTGAACATCATCGCGCGCAACATGCTCACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGGCCCCACCGCGTGAACATCATCGCGCGCAACATGCTCACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAACTTTCCCATCAGCCCCCATCGAGACCGGTGCCCCGTGAAGCTGAAAGCCCGCATG 720  
Db 661 ACCCTGAACTTTCCCATCAGCCCCCATCGAGACCGGTGCCCCGTGAAGCTGAAAGCCCGCATG 720  
QY 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCCGC 780  
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCCGC 780  
QY 781 ATCTCGAGGAGATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840  
Db 781 ATCTCGAGGAGATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCCGAGAACCCCTAC 840  
QY 841 AACACCCCGGTGTGGCCATCAAGAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGGTGTGGCCATCAAGAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
QY 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCGCCAC 960  
Db 901 TTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCGCCAC 960  
QY 961 CCCGCGGCGCTGAAGAAGAAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTACTTTC 1020  
Db 961 CCCGCGGCGCTGAAGAAGAAGAGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTACTTTC 1020  
QY 1021 AGCGTGCCCTTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCCAGCATCAAC 1080  
Db 1021 AGCGTGCCCTTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCCCAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGGCATCCGCTACAGTACAAACGTGTGCCCGGCGGTGGAGGGGAGC 1140  
Db 1081 AACGAGACCCCGGGCATCCGCTACAGTACAAACGTGTGCCCGGCGGTGGAGGGGAGC 1140  
QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCGCTTCCGCGCGCCGCAACCC 1200

```
Db 1141 CCACAGCATCTCCAGACGAGCATGACCAAGATCTGAGAGCCCTTCGCGCCGCAACCCC 1200
Qy 1201 GAGATCGTGTATCTACCAAGGCCCCCTCTAGCTGGGAGAGCACTTGAGATCGGCCAGCAC 1260
Db 1201 GAGATCGTGTATCTACCAAGGCCCCCTCTAGCTGGGAGAGCACTTGAGATCGGCCAGCAC 1260
Qy 1261 CGCGCCAAGATCGAGAGCTGGCGAAGCACTGCTGCGCTGGGGCTTCAACAACCCCGAC 1320
Db 1261 CGCGCCAAGATCGAGAGCTGGCGAAGCACTGCTGCGCTGGGGCTTCAACAACCCCGAC 1320
Qy 1321 AAGAGCACCAAGAGAGGCCCCCTCTCTGCCAT-----CGAGCTGCACCCCGACAAG 1374
Db 1321 AAGAGCACCAAGAGAGGCCCCCTCTCTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy 1375 TGACACCTGTCAGGCCCTCGAGCTGCCCGAGAAGAGAGCTGGACCGTGAACGACATCCAG 1434
Db 1381 TGGACCTGTCAGGCCCTCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy 1435 AAGCTGTTGGGCAAGCTGAATCTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1494
Db 1441 AAGCTGTTGGGCAAGCTGAATCTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy 1495 CTGTGCAAGCTGTGCGCGGCGCCNAGGCCCTGACCGACATCTGTCCTGACCGAGGAG 1554
Db 1501 CTGTGCAAGCTGTGCGCGGCGCCNAGGCCCTGACCGACATCTGTCCTGACCGAGGAG 1560
Qy 1555 GCGGAGCTGAGCTGGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGACGCGCTGTATC 1614
Db 1561 GCGGAGCTGAGCTGGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGACGCGCTGTATC 1620
Qy 1615 TACGACCCCAAGAGCACTGGTGGCGAGATCCAGAGCAGGCGCCACGACCAAGTGGACC 1674
Db 1621 TACGACCCCAAGAGCACTGGTGGCGAGATCCAGAGCAGGCGCCACGACCAAGTGGACC 1680
Qy 1675 TACGAGATCTACGAGAGCCCTTCAGAACTGAAGACCGGCAAGTACGCGCAAGATGCGC 1734
Db 1681 TACGAGATCTACGAGAGCCCTTCAGAACTGAAGACCGGCAAGTACGCGCAAGATGCGC 1740
Qy 1735 ACCGCCCAACCAAGACGCTGAAGACGCTGACGAGGCGCTGAGAGATCGCCATGGAG 1794
Db 1741 ACCGCCCAACCAAGACGCTGAAGACGCTGACGAGGCGCTGAGAGATCGCCATGGAG 1800
Qy 1795 AGCATCGTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGAGACCTGG 1854
Db 1801 AGCATCGTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGAGAGACCTGG 1860
Qy 1855 GAGACCTGGTGGACCGACTACTGCGAGGCCCACTGGATCCCGAGTGGGATTCGTGAAC 1914
Db 1861 GAGACCTGGTGGACCGACTACTGCGAGGCCCACTGGATCCCGAGTGGGATTCGTGAAC 1920
Qy 1915 ACCCCCCCTGTTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGCGCGGAG 1974
Db 1921 ACCCCCCCTGTTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGCGCGGAG 1980
Qy 1975 ACCTTCTACGTGACGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTG 2034
Db 1981 ACCTTCTACGTGACGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
Qy 2035 ACCGACCGGGGCGGCAAGATCTGTAGCTGAGCGAGACCAACCAACGAAAGACCGAG 2094
Db 2041 ACCGACCGGGGCGGCAAGATCTGTAGCTGAGCGAGACCAACCAACGAAAGACCGAG 2100
Qy 2095 CTGACGSCCATCCAGCTGGCCCTGCGAGGACAGCGGAGCGTGAACATCGTGACCGAC 2154
Db 2101 CTGACGSCCATCCAGCTGGCCCTGCGAGGACAGCGGAGCGTGAACATCGTGACCGAC 2160
Qy 2155 AGCCAGTACCCCTGGGCACTATCCAGGCCCCAGCCGACAAAGAGCGAGAGCGTGGTG 2214
Db 2161 AGCCAGTACCCCTGGGCACTATCCAGGCCCCAGCCGACAAAGAGCGAGAGCGTGGTG 2220
Qy 2215 AACGAGATCATCAGACGCTGATCAAGAGGAGAGGTGATCTGAGCTGGTGCCCGCC 2274
```

```
Db 2221 AACGAGATCATCAGACGAGCTGATCAAGAGAGAGGTGATCCTGAGCTGGTGCCCGCC 2280
Qy 2275 CACAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGGCATCCGCAAG 2334
Db 2281 CACAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGGCATCCGCAAG 2340
Qy 2335 GTCTGTCTTCGACCGGCACTGATGCGGCGCATCGTGATCTACCAAGTACATGAGACGCTG 2394
Db 2341 GTCTGTCTTCGACCGGCACTGATGCGGCGCATCGTGATCTACCAAGTACATGAGACGCTG 2400
Qy 2395 TACGTGGGAGCGCGGCCCTTAGATCGATTAAGCTTCCCGGGGCTAGACCGGT 2451
Db 2401 TACGTGGGAGCGCGGCCCTTAGATCGATTAAGCTTCCCGGGGCTAGACCGGT 2457

RESULT 14
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein; immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef; immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
XX
SY Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR ) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef (i). (1) have immunostimulant activity and can be used in gene therapy. The HIV type C polynucleotides are useful in applications including DNA immunisation, generation of packaging cell lines, and production of HIV type C proteins. The polynucleotides are particularly useful in gene therapy and DNA immunisation applications. ABL39942 to ABL40054 and CC ABB06204 to ABB06215 represent sequences used in the exemplification of the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 1.4e-290;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

Qy 1 GTGACGCCACCATGCGCGAGGCGCATGAGCCAGGCCAGGCCAACATCTCTGATGCGAG 60
```

Db 1 GTGACGCCACCATGGCGGAGGCCCATGAGCCAGGCCACACAGCGCAACATCCTGATGAG 60  
Qy 61 CGCAGCACTTCAAGGGCCCCAAGGCGATCATCAAGTCTTCAACTGGGCGCAAGAGGGC 120  
Db 61 CGCAGCACTTCAAGGGCCCCAAGGCGATCATCAAGTCTTCAACTGGGCGCAAGAGGGC 120  
Qy 121 CACATCGCCCCCAACTGCGCGCCCCCGCGAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCCCCAACTGCGCGCCCCCGCGAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180  
Qy 181 GGCACACAGATGAAGAGATGACACGAGCGCGACGCCAACTTCTTCCGCGAGGACCTGGCC 240  
Db 181 GGCACACAGATGAAGAGATGACACGAGCGCGACGCCAACTTCTTCCGCGAGGACCTGGCC 240  
Qy 241 TTCCCCCAGGGCAGGGCCCCGGAGTCCCCCAGCGAGCAGAACCGCGCCAAACAGCCCCCACC 300  
Db 241 TTCCCCCAGGGCAGGGCCCCGGAGTCCCCCAGCGAGCAGAACCGCGCCAAACAGCCCCCACC 300  
Qy 301 AGCGCGAGCTGCAAGGTGCGCGCGCAAAACCCCGCAGCGAGGGCGCGCGCGAGCGCCAG 360  
Db 301 AGCGCGAGCTGCAAGGTGCGCGCGCAAAACCCCGCAGCGAGGGCGCGCGCGAGCGCCAG 360  
Qy 361 GGCACCTGAACTTCCCCAGATCACCTGTGTGAGCGCCCCCTGTGTGAGCATCAAGGTG 420  
Db 361 GGCACCTGAACTTCCCCAGATCACCTGTGTGAGCGCCCCCTGTGTGAGCATCAAGGTG 420  
Qy 421 GCGCGCAGATCAAGGAGGGCCCTGCTGGACACCGCGCGCAGACACCGTGTGAGGAG 480  
Db 421 GCGCGCAGATCAAGGAGGGCCCTGCTGGACACCGCGCGCAGACACCGTGTGAGGAG 480  
Qy 481 ATGAGCCTGCCCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540  
Db 481 ATGAGCCTGCCCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG 540  
Qy 541 GTGCGCAGTACGACCAAGATCCTGATCGAGATCTGCGCGCAAGAGGCCATCGGCACCGTG 600  
Db 541 GTGCGCAGTACGACCAAGATCCTGATCGAGATCTGCGCGCAAGAGGCCATCGGCACCGTG 600  
Qy 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Qy 661 ACCCTGAACTTCCCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGGCATG 720  
Qy 721 GACGGCCCCAAGGTGAAGTGAAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC 780  
Db 721 GACGGCCCCAAGGTGAAGTGAAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC 780  
Qy 781 ATCTGCGAGGATGGAGAGGAGGGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGATGGAGAGGAGGGCAAGATCACCAAGATCGGCGCGAGAACCCCTAC 840  
Qy 841 AACACCCCGTGTTCGCGCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGTGTTCGCGCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCAATCCCCCACC 960  
Db 901 TTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCAATCCCCCACC 960  
Qy 961 CCGCGCGCTGAGAGAGAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTACTTC 1020  
Db 961 CCGCGCGCTGAGAGAGAGAGAGAGCGTGAACCGTGTGGACGTGGGCGAGCGCTACTTC 1020  
Qy 1021 AGCGTGGCCCTGGAGCAGGACTTCCGCAAGTACACCGCTTCACCATCCCCCAGCATCAAC 1080  
Db 1021 AGCGTGGCCCTGGAGCAGGACTTCCGCAAGTACACCGCTTCACCATCCCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGATCCCGTACACGATACAACTGTGTGCCCGCAGGCTGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGCGATCCCGTACACGATACAACTGTGTGCCCGCAGGCTGAAGGGCAGC 1140

Qy 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTTCGCGCGCCGCAACCCC 1200  
Db 1141 CCAGCATCTTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTTCGCGCGCCGCAACCCC 1200  
Qy 1201 GAGATCGTGTATCTACCA-----GGCCCCCTGTAGCTGGGCGAGCGACCTGGAGATCGGC 1254  
Db 1201 GAGATCGTGTATCTACCA-----GGCCCCCTGTAGCTGGGCGAGCGACCTGGAGATCGGC 1260  
Qy 1255 CAGCACCGCGCAAGATCGAGAGAGCTGCGAAGACCTGCTGCGCTGGGGCTTCACACCC 1314  
Db 1261 CAGCACCGCGCAAGATCGAGAGAGCTGCGAAGACCTGCTGCGCTGGGGCTTCACACCC 1320  
Qy 1315 CCGACAAGAAGACCAAGAAAGGAGCCCCCTTCTCTGCCAT-----CGAGCTGCACCCC 1368  
Db 1321 CCGACAAGAAGACCAAGAAAGGAGCCCCCTTCTCTGGATGGGCTAGAGCTGCACCCC 1380  
Qy 1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1428  
Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1440  
Qy 1429 ATCCAGAAGCTGGTGGCAAGCTGAATGGGCGACCGAGATCTACCCCGGCGATCAAGGTG 1488  
Db 1441 ATCCAGAAGCTGGTGGGCAAGCTGAATGGGCGACCGAGATCTACCCCGGCGATCAAGGTG 1500  
Qy 1489 CGCCAGCTGTCAAGCTGTGCGCGCGCAAGSCCTGACCGACATCTGTGCCCTGACC 1548  
Db 1501 CGCCAGCTGTCAAGCTGTGCGCGCGCGCAAGSCCTGACCGACATCTGTGCCCTGACC 1560  
Qy 1549 GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGGAGCCCGTGCACGGC 1608  
Db 1561 GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGGAGCCCGTGCACGGC 1620  
Qy 1609 GTGTACTACGACCCCGAGCAAGGACCTGGTGGCGGAGATCCAGAAAGAGGGCGACGACG 1668  
Db 1621 GTGTACTACGACCCCGAGCAAGGACCTGGTGGCGGAGATCCAGAAAGAGGGCGACGACG 1680  
Qy 1669 TGAACCTTACCAGATCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1728  
Db 1681 TGAACCTTACCAGATCTACCAAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAG 1740  
Qy 1729 ATGCGCACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAAGATCGCC 1788  
Db 1741 ATGCGCACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAAGATCGCC 1800  
Qy 1789 ATGAGAGCATGTGTATCTGGGCGAGACCCCGCAAGTTCGCGCTGCCCATCCAGAAAGGAG 1848  
Db 1801 ATGAGAGCATGTGTATCTGGGCGAGACCCCGCAAGTTCGCGCTGCCCATCCAGAAAGGAG 1860  
Qy 1849 ACCTGGAGACCTGGTGGACCGGACTCTGCGAGCGCACCTGGATCCCCCGAGTGGAGTTC 1908  
Db 1861 ACCTGGAGACCTGGTGGACCGGACTCTGCGAGCGCACCTGGATCCCCCGAGTGGAGTTC 1920  
Qy 1909 GTGAACACCCCGCCCTGGTGAAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCGGC 1968  
Db 1921 GTGAACACCCCGCCCTGGTGAAGCTGTGTGTACAGCTGGAGAGGAGCCCATCATCGGC 1980  
Qy 1969 GCGAGACCTTCTAGCTGGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGGCGGC 2028  
Db 1981 GCGAGACCTTCTAGCTGGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGGCGGC 2040  
Qy 2029 TAGCTGACCCAGCGGGCGGAGAGATCGTAGGCTGACCGAGACCAACCAAGAG 2088  
Db 2041 TAGCTGACCCAGCGGGCGGAGAGATCGTAGGCTGACCGAGACCAACCAAGAG 2100  
Qy 2089 ACCGAGCTGAGGCCATTCAGCTGGCCCTGACGACAGCGCGCAGCGAGGTGAACATCGTG 2148  
Db 2101 ACCGAGCTGAGGCCATTCAGCTGGCCCTGACGACAGCGCGCAGCGAGGTGAACATCGTG 2160  
Qy 2149 ACCGAGCGCAGTACGCCCTGGGCGATCATCCAGGCCAGCCCGACAGAGCGGAGCGAG 2208  
Db 2161 ACCGAGCGCAGTACGCCCTGGGCGATCATCCAGGCCAGCCCGACAGAGCGGAGCGAG 2220

Qy	2209	CTGCTGAACCAAGATCATCGAGCAGCTGATCAAGAAAGGAGGTGTACCTGAGCTGGGTG	2266
Db	2221	CTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAAGGAGGAGGTGTACCTGAGCTGGGTG	2280
Qy	2269	CCGCCCCACAAGGGCATCGCGGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCAGTC	2328
Db	2281	CCCGCCCCACAAGGGCATCGCGGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCAGTC	2340
Qy	2329	CGCAAGGTGCTGTTCTCTGGACGGGCATCGATGGCGGCATCGTGTGATCTACCAAGTACATGGAC	2388
Db	2341	CGCAAGGTGCTGTTCTCTGGACGGGCATCGATGGCGGCATCGTGTGATCTACCAAGTACATGGAC	2400
Qy	2389	GACCTGTACGTGGGACGGCGGCCCTTAGAGATCGATTAAAAGCTTCCCGGGGTAGCACC	2448
Db	2401	GACCTGTACGTGGGACGGCGGCCCTTAGAGATCGATTAAAAGCTTCCCGGGGTAGCACC	2460
Qy	2449	GGTGAATTC 2457	
Db	2461	GGTGAATTC 2469	
RESULT 15			
ADM73764			
ID	ADM73764 standard; DNA; 2469 BP.		
XX	ADM73764;		
XX			
DT	03-JUN-2004 (first entry)		
XX			
DE	HIV-1 polynucleotide #7.		
XX			
KW	HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;		
KW	HIV type C protein; immunostimulant.		
XX			
OS	Human immunodeficiency virus 1.		
XX			
PN	US2003223961-A1.		
XX			
PD	04-DEC-2003.		
XX			
PP	05-JUL-2001; 2001US-00899575.		
XX			
PR	05-JUL-2000; 2000US-00610313.		
XX			
PA	(MEGE/) MEGEDE J Z.		
PA	(BARN/) BARNETT S W.		
PA	(ENGE/) ENGELBRECHT S.		
PA	(RENS/) RENSBERG E J V.		
XX			
PI	Megede JZ., Barnett SW, Engelbrecht S, Rensburg EJV;		
XX			
DR	WPI; 2004-060515/06.		
XX			
PT	New expression cassette comprising a polynucleotide sequence encoding an		
PT	HIV Pol polypeptide, useful in eliciting an immune response, in DNA		
PT	immunization, generating of packaging cell lines or in producing HIV Type		
PT	C proteins.		
XX			
PS	Claim 1; SEQ ID NO 30; 160pp; English.		
XX			
CC	The invention relates to an expression cassette comprising a		
CC	polynucleotide sequence encoding an HIV Pol polypeptide. The invention		
CC	also relates to a recombinant expression system for use in a host cell		
CC	comprising an expression cassette, where the polynucleotide sequence		
CC	further comprises control elements capable of driving expression in the		
CC	selected host cell, a cell comprising an expression cassette where the		
CC	polynucleotide sequence further comprises control elements compatible		
CC	with the expression in the cell and a composition for generating an		
CC	immunological response, comprising an expression cassette. The expression		
CC	cassette and the methods of the invention are useful in eliciting an		
CC	immune response, in DNA immunisation, in generation of packaging cell		
CC	lines and in producing HIV Type C proteins. This sequence represents an		
CC	HIV-1 polynucleotide of the invention.		

Db 961 CCCCGCGGCTCAAGAGAAGAGCGTGCACCGTGTGGACGTGGCGAGCGCTACTTTC 1020  
 QY 1021 ACCGTGCCCCCTGGACGAGACTTTCGCGAAGTACACCGCTTACCATCCCGAGATCAAC 1080  
 Db 1021 AGCGTGCCTTGGACGAGGACTTTCGCGAAGTACACCGCTTACCATCCCGAGATCAAC 1080  
 QY 1081 AACGAGACCCCGCGCATCCGCTACAGTACAGTACAGTGCCTCCCGAGGCTTGGAGGGCAGC 1140  
 Db 1081 AACGAGACCCCGCGCATCCGCTACAGTACAGTACAGTGCCTCCCGAGGCTTGGAGGGCAGC 1140  
 QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGGCGCCGCAACCCC 1200  
 Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGGCGCCGCAACCCC 1200  
 QY 1201 GAGATCGTGATACCA-----GGCCCCCTGTACGTGGGAGCGACCTTGGAGATCGGC 1254  
 Db 1201 GAGATCGTGATACCAAGCATGATGACGACCTGTACGTGGGAGCGACCTTGGAGATCGGC 1260  
 QY 1255 CAGCACCGGCCAAGATCGAGGAGCTGCGACGACCTGCTGCGCTGGGCTTCAACACC 1314  
 Db 1261 CAGCACCGGCCAAGATCGAGGAGCTGCGACGACCTGCTGCGCTGGGCTTCAACACC 1320  
 QY 1315 CCCGACAAGACACAGAAAGAGCCCTTCTCTGCGCAT-----CGAGCTGCACCCC 1368  
 Db 1321 CCCGACAAGACACAGAAAGAGCCCTTCTCTGCGCAT-----CGAGCTGCACCCC 1380  
 QY 1369 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAAGAGAGCTGGACCGTGAACGAC 1428  
 Db 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAAGAGAGCTGGACCGTGAACGAC 1440  
 QY 1429 ATCCAGAAAGCTGGTGGGCAAGCTGAATCGGGCCAGCCAGATCTACCCGGGCATCAAGGTG 1488  
 Db 1441 ATCCAGAAAGCTGGTGGGCAAGCTGAATCGGGCCAGCCAGATCTACCCGGGCATCAAGGTG 1500  
 QY 1489 CCCAGAGCTGTGAAGCTGTGCGCGCGCCAGAGCCCTGACCGACATCGTGCCTCGACC 1548  
 Db 1501 CGCCAGCTGTGAAGCTGTGCGCGCGCCAGAGCCCTGACCGACATCGTGCCTCGACC 1560  
 QY 1549 GAGGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGATCTTGCAGGAGCCCGTGCACGGC 1608  
 Db 1561 GAGGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGATCTTGCAGGAGCCCGTGCACGGC 1620  
 QY 1609 GTGTACTAGACCCAGCAAGACCTGTGTGGCCGAGATCCAGAACGAGGCCACGACCCAG 1668  
 Db 1621 GTGTACTAGACCCAGCAAGACCTGTGTGGCCGAGATCCAGAACGAGGCCACGACCCAG 1680  
 QY 1669 TGGACCTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1728  
 Db 1681 TGGACCTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1740  
 QY 1729 ATGCGGACCGCCCAACCAAGCTGAAGCAGCTGACCGAGGCGGTGGAGAGATCGCC 1788  
 Db 1741 ATGCGGACCGCCCAACCAAGCTGAAGCAGCTGACCGAGGCGGTGGAGAGATCGCC 1800  
 QY 1789 ATGGAGAGCATCGTGAATCTGGGGCAGACCCCGAAGTTCGCGCTGCCCATCCAGAGGAG 1848  
 Db 1801 ATGGAGAGCATCGTGAATCTGGGGCAGACCCCGAAGTTCGCGCTGCCCATCCAGAGGAG 1860  
 QY 1849 ACCTGGGAGACCTGTGGACCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTTC 1908  
 Db 1861 ACCTGGGAGACCTGTGGACCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTTC 1920  
 QY 1909 GTGAACACCCCCCTTGGTGAAGCTGTGGTACAGCTGGAGAAAGGAGCCCATCATCGGC 1968  
 Db 1921 GTGAACACCCCCCTTGGTGAAGCTGTGGTACAGCTGGAGAAAGGAGCCCATCATCGGC 1980  
 QY 1969 GCGGAGACCTTCTAGTGGACCGCGCCGACCGGAGACCAAGATCGGCAAGGCCGCGC 2028  
 Db 1981 GCGGAGACCTTCTAGTGGACCGCGCGCCGACCGGAGACCAAGATCGGCAAGGCCGCGC 2040  
 QY 2029 TACGTGACCGACCGGGGCGCGCAGAGATCGTGAAGCTGTGAGCTGACCGAGACCAACCAAGAG 2088  
 Db 2041 TACGTGACCGACCGGGGCGCGCAGAGATCGTGAAGCTGTGAGCTGACCGAGACCAACCAAGAG 2100

QY 2089 ACCGAGCTGCAGGCCATCCAGCTGGCCTTGAGGACAGCGGAGCGAGGTGAACATCGTG 2148  
 Db 2101 ACCGAGCTGCAGGCCATCCAGCTGGCCTTGAGGACAGCGGAGCGAGGTGAACATCGTG 2160  
 QY 2149 ACCGACAGCCAGTACGCCCTTGGGCATCATCCAGGCCAGCCGACCAAGAGGAGAGCGAG 2208  
 Db 2161 ACCGACAGCCAGTACGCCCTTGGGCATCATCCAGGCCAGCCGACCAAGAGGAGAGCGAG 2220  
 QY 2209 CTGCTGAACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAAAGGTGTACTCTGAGCTGGGTG 2268  
 Db 2221 CTGCTGAACCCAGATCATCGAGCAGCTGATCAAGAAAGGAGAAAGGTGTACTCTGAGCTGGGTG 2280  
 QY 2269 CCGCCCCCAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGGTGCAGCAAGGGCATC 2328  
 Db 2281 CCGCCCCCAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGGTGCAGCAAGGGCATC 2340  
 QY 2329 CGCAAGGTGCTTCTCTGAGCGGATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2388  
 Db 2341 CGCAAGGTGCTTCTCTGAGCGGATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2400  
 QY 2389 GACCTGTACGTGGGCGAGCGGCGCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACC 2448  
 Db 2401 GACCTGTACGTGGGCGAGCGGCGCTTAGGATCGATTAAGCTTCCCGGGGCTAGCACC 2460  
 QY 2449 GGTGAATTC 2457  
 Db 2461 GGTGAATTC 2469

Search completed: June 1, 2005, 11:33:34  
 Job time : 1272.25 secs





QY 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGGGCAAGAGGCG 120  
DB 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGGGCAAGAGGCG 120  
QY 121 CACATCGCCGCAACTGCGGCGCCCCCGCAAGAGGCTGCTGGAAGTGGGCAAGAG 180  
DB 121 CACATCGCCGCAACTGCGGCGCCCCCGCAAGAGGCTGCTGGAAGTGGGCAAGAG 180  
QY 181 GGCACACAGATGAAGGATGACACGAGCGCCAGGCCAACTTCTTCCGCGAGCACTGGCC 240  
DB 181 GGCACACAGATGAAGGATGACACGAGCGCCAGGCCAACTTCTTCCGCGAGCACTGGCC 240  
QY 241 TTCCGCCAGGGCAAGGCCCGCGAGTTCCACGAGCAGAAACCGCGCCAAACAGCCCCAC 300  
DB 241 TTCCGCCAGGGCAAGGCCCGCGAGTTCCACGAGCAGAAACCGCGCCAAACAGCCCCAC 300  
QY 301 AGCCGCGAGCTGCAGGTGGGCGGCAAAACCCCGCAGCGAGGCCGCGGCCAGGCCAG 360  
DB 301 AGCCGCGAGCTGCAGGTGGGCGGCAAAACCCCGCAGCGAGGCCGCGGCCAGGCCAG 360  
QY 361 GGCACCTGMACTTCCCCAGATCACCTGTGGAGCGCCCTGTGTGAGCATCAAGTG 420  
DB 361 GGCACCTGMACTTCCCCAGATCACCTGTGTGGAGCGCCCTGTGTGAGCATCAAGTG 420  
QY 421 GCGCGCCAGATCAAGAGGCCCTGCTGGACACCGCGCGCAGCACACGTCGTGTGAGGAG 480  
DB 421 GCGCGCCAGATCAAGAGGCCCTGCTGTGACACCGCGCGCAGCACACGTCGTGTGAGGAG 480  
QY 481 ATGAGCCTGCCGCGAAGTGGAGGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAG 540  
DB 481 ATGAGCCTGCCGCGAAGTGGAGGCCCAAGATGATCGGCGGCATCGGCGGCTTCAATCAAG 540  
QY 541 GTGCGCCAGTAGCAGATCCTGATCGAGATCTCGGCGAAGAGGCCATCGGCAACG 600  
DB 541 GTGCGCCAGTAGCAGATCCTGATCGAGATCTCGGCGAAGAGGCCATCGGCAACG 600  
QY 601 CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
DB 601 CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAACTTCCCATCAGCCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
DB 661 ACCCTGAACTTCCCATCAGCCCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720  
QY 721 GAGCGCCCCAGGTGACGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
DB 721 GAGCGCCCCAGGTGAAACGAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
QY 781 ATCTCGGAGGAGATGAGAGAGGCGCAAGATCACCAAGATCGGCCCCGAGAAACCCCTAC 840  
DB 781 ATCTCGGAGGAGATGAGAGAGGCGCAAGATCACCAAGATCGGCCCCGAGAAACCCCTAC 840  
QY 841 AACACCCCCGTTCCGCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAC 900  
DB 841 AACACCCCCGTTCCGCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAC 900  
QY 901 TTCGCGAGCTGAACAGCGCACCCAGACTTCTGGAGGTGACCTGGGCACTCCCCAC 960  
DB 901 TTCGCGAGCTGAACAGCGCACCCAGACTTCTGGAGGTGACCTGGGCACTCCCCAC 960  
QY 961 CCGCGCGGCTCAAGAGAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACTTTC 1020  
DB 961 CCGCGCGGCTCAAGAGAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACTTTC 1020  
QY 1021 AGCGTGCCCCCTGGAGGAGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCATCAAC 1080  
DB 1021 AGCGTGCCCCCTGGAGGAGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCATCAAC 1080  
QY 1081 AACGAGACCCCCGGGATCCGCTACCAAGTACACGTCGTGCTGCCCCAGGCTGGAGGGCAGC 1140  
DB 1081 AACGAGACCCCCGGGATCCGCTACCAAGTACACGTCGTGCTGCCCCAGGCTGGAGGGCAGC 1140  
QY 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCCGCAACCCC 1200

DB 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTCCGCGCCCGCAACCCC 1200  
QY 1201 GAGATCGTGATCTACAGGCCCCCTCTGTACTGTTGGGACGAGACTTGGAGATCGGCCAGCAC 1260  
DB 1201 GAGATCGTGATCTACAGGCCCCCTCTGTACTGTTGGGACGAGACTTGGAGATCGGCCAGCAC 1260  
QY 1261 CGGCCAAGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCAACACCCCGAC 1320  
DB 1261 CGGCCAAGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTCAACACCCCGAC 1320  
QY 1321 AAGAAGCAACCAAGAGGAGCCCCCTTCTGCCCATCGAGCTGCACCCCGCAAGTGGACC 1380  
DB 1321 AAGAAGCAACCAAGAGGAGCCCCCTTCTGCCCATCGAGCTGCACCCCGCAAGTGGACC 1380  
QY 1381 GTGACGCCATCGAGCTGCCGAGAGGAGCTGGAACGTTGAACGACATCCAGAGCTG 1440  
DB 1381 GTGACGCCATCGAGCTGCCGAGAGGAGCTGGAACGTTGAACGACATCCAGAGCTG 1440  
QY 1441 GTGGCAAGCTGMACTGGGCGCAGCCAGATCTACCCCGGCATCAAGTGGCCAGCTGTGC 1500  
DB 1441 GTGGCAAGCTGMACTGGGCGCAGCCAGATCTACCCCGGCATCAAGTGGCCAGCTGTGC 1500  
QY 1501 AAGCTGCTGCGCGCGCCCAAGGCCCTTGACGACATCTGTCGCCCTGACCGAGGAGCCGAG 1560  
DB 1501 AAGCTGCTGCGCGCGCCCAAGGCCCTTGACGACATCTGTCGCCCTGACCGAGGAGCCGAG 1560  
QY 1561 CTGAGCTGGCGCGAGAACCGCGAGATCTGTCGCGAGCCCGTGCAACGCGGTGTACTACGAC 1620  
DB 1561 CTGAGCTGGCGCGAGAACCGCGAGATCTGTCGCGAGCCCGTGCAACGCGGTGTACTACGAC 1620  
QY 1621 CCCAGAGGAGCTGTTGGCGCGAGATCCAGAGCAGGCGCCAGCAGCTGGAGCTTACCAG 1680  
DB 1621 CCCAGAGGAGCTGTTGGCGCGAGATCCAGAGCAGGCGCCAGCAGCTGGAGCTTACCAG 1680  
QY 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAGTGCACACCGCC 1740  
DB 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCCAAGTGCACACCGCC 1740  
QY 1741 CACACACAGAGCTGAAGCAGCTGACCGAGCGCTGCAAGAGATCGCCATGGAGAGCATC 1800  
DB 1741 CACACACAGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATC 1800  
QY 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACTTGGAGAGCC 1860  
DB 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACTTGGAGAGCC 1860  
QY 1861 TGGTGACCCGACTACTGGCAGGCCACTTGGATCCCCGAGTGGGAGTTGGTGAACACCCC 1920  
DB 1861 TGGTGACCCGACTACTGGCAGGCCACTTGGATCCCCGAGTGGGAGTTGGTGAACACCCC 1920  
QY 1921 CCCCTGGTGAAGCTGTGGTACAGCTGGAGAGAGGCCATCATCGGCGCCGAGACCTTTC 1980  
DB 1921 CCCCTGGTGAAGCTGTGGTACAGCTGGAGAGAGGCCATCATCGGCGCCGAGACCTTTC 1980  
QY 1981 TACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTTACGTGACCCAG 2040  
DB 1981 TACGTGGAGCGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTTACGTGACCCAG 2040  
QY 2041 CGGCGCCGCGCAGAAATCGTGAGCTGACCGAGACCAACCAAGAGAGCCGAGCTGACAG 2100  
DB 2041 CGGCGCCGCGCAGAAATCGTGAGCTGACCGAGACCAACCAAGAGAGCCGAGCTGACAG 2100  
QY 2101 GGCATCAGCTGGCCCTGAGAGACAGCGCGCAGAGGTAACATCGTGACCCAGAGCCAG 2160  
DB 2101 GGCATCAGCTGGCCCTGAGAGACAGCGCGCAGAGGTAACATCGTGACCCAGAGCCAG 2160  
QY 2161 TACGCCCTGGGCTCATCATCAGGCCAGCCCGCAAGAGCGAGCGAGCTGTGTGAACCCAG 2220  
DB 2161 TACGCCCTGGGCTCATCATCAGGCCAGCCCGCAAGAGCGAGCGAGCTGTGTGAACCCAG 2220  
QY 2221 ATCATCGAGCAGCTGATCAAGAGGAGAGGTTGTAACCTGAGCTGGGTGCCCGCCCAACAAG 2280

```
Db 2221 ATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTGAGCTGGTGCCCGCCCAACAG 2280
Qy 2281 GGCATCGCGCGCAACGACAGATGACAAAGCTGGTGAGCAAGGGGCATCCGCAAGTGTCTG 2340
Db 2281 GGCATCGCGCGCAACGACAGATGACAAAGCTGGTGAGCAAGGGGCATCCGCAAGTGTCTG 2340
Qy 2341 TTCCTGAGCGCATGATCGCGGCATCGTGATCTACCGATACATGAGCAGCCTGTACGTG 2400
Db 2341 TTCCTGAGCGCATGATCGCGGCATCGTGATCTACCGATACATGAGCAGCCTGTACGTG 2400
Qy 2401 GGCAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457
Db 2401 GGCAGCGCGCGCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457

RESULT 2
US-10-190-435-43
; Sequence 43, Application US/10190435
; Publication No. US20030143248A1
; GENERAL INFORMATION:
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan W.
; APPLICANT: LIAN, Ying
; APPLICANT: ENGELBRECHT, Susan
; APPLICANT: VAN RENSBERG, Esclrelita J.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; FILE REFERENCE: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP18133.003 / 2302-18133
; CURRENT APPLICATION NUMBER: US/10/190.435
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 319
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 43
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: p2Pol.opt.YMMW_C
US-10-190-435-43

Query Match 99.4%; Score 2443.4; DB 15; Length 2445;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 7 GCCACCATGGCCGAGGCGCATGAGCGAGCGCCACGAGCGCCAAACATCTCTGATGCGAGCG 66
Db 1 GCCACCATGGCCGAGGCGCATGAGCGAGCGCCACGAGCGCCAAACATCTCTGATGCGAGCG 60
Qy 67 AACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGGAGGCGCAATC 126
Db 61 AACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGGAGGCGCAATC 120
Qy 127 GCCCGCAACTGCGCGCGCCCGCCGCAAGAGGGCTCTGGAGTGGCGCAAGGAGGGCCAC 186
Db 121 GCCCGCAACTGCGCGCGCCCGCCGCAAGAGGGCTCTGGAGTGGCGCAAGGAGGGCCAC 180
Qy 187 CAGATGAAGGACTGCAACGAGCGCGCAGCGCCAACTTTCTTCCGCGAGGACCTGGCTTCCC 246
Db 181 CAGATGAAGGACTGCAACGAGCGCGCAGCGCCAACTTTCTTCCGCGAGGACCTGGCTTCCC 240
Qy 247 CAGGCGAAGCGCCCGAGTTCCTCCGAGCGAGCAAGAACCGCGCAACAGCCCGCCACAGCGCG 306
Db 241 CAGGCGAAGCGCCCGAGTTCCTCCGAGCGAGCAAGAACCGCGCAACAGCCCGCCACAGCGCG 300
Qy 307 GAGCTGAGGTGGCGGGGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGAGCGCGAGGGCAAC 366
Db 301 GAGCTGAGGTGGCGGGGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGAGCGCGAGGGCAAC 360
Qy 367 CTGAACCTTCCCGCAGATCACCTCTGTGCGAGCGCGCCCTGTTGAGCATCAAGGTGGCGGG 426
Db 361 CTGAACCTTCCCGCAGATCACCTCTGTGCGAGCGCGCCCTGTTGAGCATCAAGGTGGCGGG 420
Qy 427 CAGATCAAGGAGGGCCCTGCTGGACACCGGGCGCGAGCACACCGTGTGGAGGAGATGAGC 486
```

```
Db 421 CAGATCAAGGAGGGCCCTGCTGGCCACCGGGCCGACGACACCGTGTGGAGGAGATGAGC 480
Qy 487 CTGCCCCGGAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGCTTTCATCAAGGTGGCG 546
Db 481 CTGCCCCGGAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGCTTTCATCAAGGTGGCG 540
Qy 547 CAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 606
Db 541 CAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGATC 600
Qy 607 GGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGTCACCGCTG 666
Db 601 GGGCCCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGTCACCGCTG 660
Qy 667 AACTTTCCTCCATCAGCCCCCATCGAGACCGTGCCTGAAAGCTGGAAGCCCGGCATCGAGCGC 726
Db 661 AACTTTCCTCCATCAGCCCCCATCGAGACCGTGCCTGAAAGCTGGAAGCCCGGCATCGAGCGC 720
Qy 727 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGC 780
Qy 787 GAGGAGATGGAAGAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAACC 846
Db 781 GAGGAGATGGAAGAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTCAACAACC 840
Qy 847 CCGTGTTCCTCCATCAAGAAAGAGGACAGCACCAGGTGGCGCAAGCTGTGGACTTCGCG 906
Db 841 CCGTGTTCCTCCATCAAGAAAGAGGACAGCACCAGGTGGCGCAAGCTGTGGACTTCGCG 900
Qy 907 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCCCACCCTCGGCC 966
Db 901 GAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCCCACCCTCGGCC 960
Qy 967 GGCCTGAAGAAAGAAAGAGCGGTGAACCGTGTGAGACGTGGGCGAGCGCTACTTTCAGCGTG 1026
Db 961 GGCCTGAAGAAAGAAAGAGCGGTGAACCGTGTGAGACGTGGGCGAGCGCTACTTTCAGCGTG 1020
Qy 1027 CCCTTGACCGAGGACTTCCCGCAAGTACACCGCTTCCACCATCCCCAGCATCAACAACGAG 1086
Db 1021 CCCTTGACCGAGGACTTCCCGCAAGTACACCGCTTCCACCATCCCCAGCATCAACAACGAG 1080
Qy 1087 ACCCCCCGCGATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1146
Db 1081 ACCCCCCGCGATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGGCTGGAAGGGCAGCCCCAGC 1140
Qy 1147 ATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCGCTTTCGCGCGCCGGAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGCATGACCAAGATCTCTGAGCGCTTTCGCGCGCCGGAACCCCGAGATC 1200
Qy 1207 GTGATCTACCAAGCGCCCTGTAGTGGGCGAGCGACCTGGAGATCGGCGCAGCGCGGCC 1266
Db 1201 GTGATCTACCAAGCGCCCTGTAGTGGGCGAGCGACCTGGAGATCGGCGCAGCGCGGCC 1260
Qy 1267 AAGATCGAGGAGCTGGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCCGCAACAAGAG 1326
Db 1261 AAGATCGAGGAGCTGGCGCAAGCACCTGCTGCGTGGGGCTTTCACACCCCCGCAACAAGAG 1320
Qy 1327 CACCAAGAGGAGCGCCCTTCTGCGCATCGAGCTGCAACCCCGCAACAAGTGGAGCGTGCAG 1386
Db 1321 CACCAAGAGGAGCGCCCTTCTGCGCATCGAGCTGCAACCCCGCAACAAGTGGAGCGTGCAG 1380
Qy 1387 CCATCGAGCTGGCGAGGAGAGCTGACCGTGAACCATCCAGAGAGCTGGTGGGC 1446
Db 1381 CCATCGAGCTGGCGAGGAGAGCTGACCGTGAACCATCCAGAGAGCTGGTGGGC 1440
Qy 1447 AAGCTGAACCTGGCGCAGCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGCAAGCTG 1506
Db 1441 AAGCTGAACCTGGCGCAGCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGCAAGCTG 1500
Qy 1507 CTGCGGCGCCCAAGGCGCTGACCGCATCGTGCCTGACCGAGGAGCGCGAGCTGGAG 1566
```

```

Db 1501 CTGCGCGGCCCAAGCCCTGACCGACATCGTGCCCTGACCGAGGAGCGAGCTGGAG 1560
QY 1567 CTGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGACCGGCTGTACTACGACCCGAGC 1626
Db 1561 CTGCGCGAGAACCGGAGATCTGCGCGAGCCCGTGACCGGCTGTACTACGACCCGAGC 1620
QY 1627 AAGGACCTGGTGGCGGAGATCCAGAAAGCAGGAGGCGCCACGACGACGACCTTACAGATCTAC 1686
Db 1621 AAGGACCTGGTGGCGGAGATCCAGAAAGCAGGAGGCGCCACGACGACGACCTTACAGATCTAC 1680
QY 1687 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCACCGCCACACCC 1746
Db 1681 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCACCGCCACACCC 1740
QY 1747 AACGACGTGAAGCAGCTGACCGAGCGCGTGCGAAGATCGCCATCGAGAGCATCTGTATC 1806
Db 1741 AACGACGTGAAGCAGCTGACCGAGCGCGTGCGAAGATCGCCATCGAGAGCATCTGTATC 1800
QY 1807 TGGGCGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGGAGACCTGGGAGACCTGGTGG 1866
Db 1801 TGGGCGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGGAGACCTGGGAGACCTGGTGG 1860
QY 1867 ACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGAGTTTGGTGAACACACCCCGCCCTG 1926
Db 1861 ACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGAGTTTGGTGAACACACCCCGCCCTG 1920
QY 1927 GTGAAGCTGTGTATCAGCTGAGAAAGAGGCCCATCATCGCGCCGAGACCTTCTACGTG 1986
Db 1921 GTGAAGCTGTGTATCAGCTGAGAAAGAGGCCCATCATCGCGCCGAGACCTTCTACGTG 1980
QY 1987 GACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGTACGTGACCGACCGGGGC 2046
Db 1981 GACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGTACGTGACCGACCGGGGC 2040
QY 2047 CGGCGAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGAGCCATC 2106
Db 2041 CGGCGAAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGAGCCATC 2100
QY 2107 GAGCTGGCCCTGACGACAGCGCGAGCGAGGTGAACATCGTGACCGACGACGATACGCC 2166
Db 2101 GAGCTGGCCCTGACGACAGCGCGAGCGAGGTGAACATCGTGACCGACGACGATACGCC 2160
QY 2167 CTGGGCATCATCAGCGCCAGCCGACAGAGCGAGCGAGCTGGTGAACAGATCATC 2226
Db 2161 CTGGGCATCATCAGCGCCAGCCGACAGAGCGAGCGAGCTGGTGAACAGATCATC 2220
QY 2227 GAGCAGCTGATCAAGAAGAGAGGTGTACTGAGCTGGGTGCCCGCCCAAGAGGCGATC 2286
Db 2221 GAGCAGCTGATCAAGAAGAGAGGTGTACTGAGCTGGGTGCCCGCCCAAGAGGCGATC 2280
QY 2287 GCGCGCAAGCAGCAGATCGAAGCTGGTGAGCAAGGGGATCCGCAAGGTGCTGTCTGT 2346
Db 2281 GCGCGCAAGCAGCAGATCGAAGCTGGTGAGCAAGGGGATCCGCAAGGTGCTGTCTGT 2340
QY 2347 GACGCGATCGATGGCGGATCGTATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 2406
Db 2341 GACGCGATCGATGGCGGATCGTATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 2400
QY 2407 GCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2451
Db 2401 GCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2445

```

RESULT 3

```

US-10-190-305A-37
; Sequence 37, Application US/10190305A
; Publication No. US20030198621A1
; GENERAL INFORMATION:
; APPLICANT: EUR MEGEDE, Jan
; APPLICANT: BARNETT, Susan
; APPLICANT: LIAN, Ying
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

```

```

; FILE REFERENCE: 2302-18702 / 18702.002
; CURRENT APPLICATION NUMBER: US/10/190.305A
; CURRENT FILING DATE: 2002-07-05
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
; LENGTH: 2445
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: p2Pol.opt.YMMW_C
US-10-190-305A-37

```

```

Query Match 99.4%; Score 2443.4; DB 16; Length 2445;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 7 GCCACCATGGCGGAGGCCATGAGCCAGGCCACCGAGCGCAACATCCTGTATGAGGCGCAGC 66
Db 1 GCCACCATGGCGGAGGCCATGAGCCAGGCCACCGAGCGCAACATCCTGTATGAGGCGCAGC 60
QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGCCACATC 126
Db 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCAAGGAGGGCCACATC 120
QY 127 GCCCGCAACTGCGCGGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGCCAC 186
Db 121 GCCCGCAACTGCGCGGCCCGCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGCCAC 180
QY 187 CAGATGAAGGACTGCAACGAGCGCGCGAGCCAACTTCTTCCGCGAGGACTGTGCTTCC 246
Db 181 CAGATGAAGGACTGCAACGAGCGCGCGAGCCAACTTCTTCCGCGAGGACTGTGCTTCC 240
QY 247 CAGGCAAGGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 306
Db 241 CAGGCAAGGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 300
QY 307 GAGCTGAGGTGGCGGGCGCAACCCCGCGAGCGAGGGCGGGCGCGAGGGCGCGAGGGCACC 366
Db 301 GAGCTGAGGTGGCGGGCGCAACCCCGCGAGCGAGGGCGGGCGCGAGGGCGCGAGGGCACC 360
QY 367 CTGAATCTCCCGCAGATCAGCTGTGGCAGCGCCCGCTGGTGAGCATCAAGGTGGCGCGC 426
Db 361 CTGAATCTCCCGCAGATCAGCTGTGGCAGCGCCCGCTGGTGAGCATCAAGGTGGCGCGC 420
QY 427 CAGATCAAGGAGGGCCCTGCTGGACACCGCGCGCGACGACACCGTGTCTGGAGGAGATGAGC 486
Db 421 CAGATCAAGGAGGGCCCTGCTGGACACCGCGCGCGACGACACCGTGTCTGGAGGAGATGAGC 480
QY 487 CTGCGCGCAAGTGGAGGCCCAAGATGATCGCGCGCATCGCGGGCTTATCAAGGTGGCG 546
Db 481 CTGCGCGCAAGTGGAGGCCCAAGATGATGATCGCGCGCATCGCGGGCTTATCAAGGTGGCG 540
QY 547 CAGTACGACCAAGATCTCTATCGAGATCTGCGCAAGAGGCCCATCGGCAACCGTGTGTATC 606
Db 541 CAGTACGACCAAGATCTCTATCGAGATCTGCGCAAGAGGCCCATCGGCAACCGTGTGTATC 600
QY 607 GCGCCCAACCCCGTGAACATCATCGCGCGCAACATGTCTGACCCAGCTGGGTGTGACCCCTG 666
Db 601 GCGCCCAACCCCGTGAACATCATCGCGCGCAACATGTCTGACCCAGCTGGGTGTGACCCCTG 660
QY 667 AACTTCCCGCATAGCCCGCATCGAGCCCGTGGTGAAGCTGGAAGCCCGCGCATGAGCGCGC 726
Db 661 AACTTCCCGCATAGCCCGCATCGAGCCCGTGGTGAAGCTGGAAGCCCGCGCATGAGCGCGC 720
QY 727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCCATCTGC 786
Db 721 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCCCATCTGC 780
QY 787 GAGGAGATGGAGAGGAGGGCAAGATCAACAGATCGCGCCCGCGAGAACCCCTTACAAACCC 846
Db 781 GAGGAGATGGAGAGGAGGGCAAGATCAACAGATCGCGCCCGCGAGAACCCCTTACAAACCC 840

```



QY	134	ACTGCGCGCCCCCGCAGAAAGGGCTGCTGGAAGTGGCGCAAGGAGGGCCACCAAGATGA	193
Db	1607	ACTGCGCGCCCCCGCAGAAAGGGTGTCTGGAAGTGGCGCAAGGAGGGCCACCAAGATGA	1666
QY	194	AGGACTGCACCGAGGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCCTTCCCGCCAGGGCA	253
Db	1667	AGGACTGCACCGAGGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCCTTCCCGCCAGGGCA	1726
QY	254	AGGCGCGAGTTCCCGAGCGAGCAACCCGCGCAACAGCGCCCAACAGCGCGCGAGCTGC	313
Db	1727	AGGCGCGAGTTCCCGAGCGAGCAACCCGCGCAACAGCGCCCAACAGCGCGCGAGCTGC	1786
QY	314	AGGTGGCGGCGACAACCCCGCAGCGAGGCGCGCGCGAGCGCGCAGGGCACCTGAACT	373
Db	1787	AGGTGGCGGCGACAACCCCGCAGCGAGGCGCGCGCGAGCGCGCAGGGCACCTGAACT	1846
QY	374	TCCCCAGATCACTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGCGCGCCAGATCA	433
Db	1847	TCCCCAGATCACTGTGGCAGCGCCCCCTGGTGAGCATCAAGGTGGCGCGCCAGATCA	1906
QY	434	AGGAGCCCTGTGACACCGCGCGCGAGCAACCGTGTCTGAGAGAGATGAGCCTGCCG	493
Db	1907	AGGAGCCCTGTGTGACACCGCGCGCGAGCAACCGTGTCTGAGAGAGATGAGCCTGCCG	1966
QY	494	GCAAGTGGAAAGCCCAAGATGATCGCGGCATCGCGGGCTTCATCAAGGTGGCCAGTAGG	553
Db	1967	GCAAGTGGAAAGCCCAAGATGATCGCGGCATCGCGGGCTTCATCAAGGTGGCCAGTAGG	2026
QY	554	ACCAGATCCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA	613
Db	2027	ACCAGATCCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA	2086
QY	614	CCCCCGTGAAATCATTCGGCCCGCAATGCTGACCCAGCTGGGCTGCAACCTGAACTTCC	673
Db	2087	CCCCCGTGAAATCATTCGGCCCGCAATGCTGACCCAGCTGGGCTGCAACCTGAACTTCC	2146
QY	674	CCATCAGGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGATGAGCGCCCAAGG	733
Db	2147	CCATCAGGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGATGAGCGCCCAAGG	2206
QY	734	TGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGCCCTGACCGCCATCTCGAGAGGA	793
Db	2207	TGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGCCCTGACCGCCATCTCGAGAGGA	2266
QY	794	TGGAGAGAGGGGCAAGATCAACAAGATCGGCCCGCGAGAAACCCCTACAAACCCCCCGTGT	853
Db	2267	TGGAGAGAGGGGCAAGATCAACAAGATCGGCCCGCGAGAAACCCCTACAAACCCCCCGTGT	2326
QY	854	TGCGCATCAAGAAGAGGACAGCAACAGTGGCGCAAGCTGGTGACCTCCGCGAGCTGA	913
Db	2327	TGCGCATCAAGAAGAGGACAGCAACAGTGGCGCAAGCTGGTGACCTCCGCGAGCTGA	2386
QY	914	ACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGATCCCCCAGCCCGCGCCCTGA	973
Db	2387	ACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGATCCCCCAGCCCGCGCCCTGA	2446
QY	974	AGAAGAAGAAGCGTGACCGTGTGGAGCGTGGGCGACGCCTACTTCAAGCGTGGCCCTGG	1033
Db	2447	AGAAGAAGAAGCGTGACCGTGTGGAGCGTGGGCGACGCCTACTTCAAGCGTGGCCCTGG	2506
QY	1034	ACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCG	1093
Db	2507	ACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCG	2566
QY	1094	GCATCCGCTACAGTACAACTGCTGCCCGAGGCTGGAAGGGCAGCGCCCGAGCATCTTCC	1153
Db	2567	GCATCCGCTACAGTACAACTGCTGCCCGAGGCTGGAAGGGCAGCGCCCGAGCATCTTCC	2626
QY	1154	AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCTGTGATCT	1213
Db	2627	AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCTGTGATCT	2686

QY	1214	ACCAGGCCCCCTGTATCGTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCAAGATCG	1273
Db	2687	ACCAGGCCCCCTGTATCGTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCAAGATCG	2746
QY	1274	AGGAGCTCGCAAGCACCTGTGCGCTGGGGCTTACACACCCCGCAAGAAGACCAAGA	1333
Db	2747	AGGAGCTCGCAAGCACCTGTGCGCTGGGGCTTACACACCCCGCAAGAAGACCAAGA	2806
QY	1334	AGGAGCCCCCTTCTCTGCCCATCGAGCTGACCCCGCAACAAGTGGACCGTGCAGCCCCATCG	1393
Db	2807	AGGAGCCCCCTTCTCTGCCCATCGAGCTGACCCCGCAACAAGTGGACCGTGCAGCCCCATCG	2866
QY	1394	AGCTGCCGAGAAAGGAGAGCTGGACCGGTGAACGACATCAGAAAGCTGTGGGCAAGCTGA	1453
Db	2867	AGCTGCCGAGAAAGGAGAGCTGGACCGGTGAACGACATCAGAAAGCTGTGGGCAAGCTGA	2926
QY	1454	ACTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCG	1513
Db	2927	ACTGGGCGAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCG	2986
QY	1514	GGGCCAAGGCCCTGACCGACATCGTGCCCCCTGACCGAGAGGCGCGAGCTGGAGCTGGCG	1573
Db	2987	GGGCCAAGGCCCTGACCGACATCGTGCCCCCTGACCGAGAGGCGCGAGCTGGAGCTGGCG	3046
QY	1574	AGAACCGGAGATCCTGCGCGAGCCGCTGACCGCGGTGTACTACGACCCCGCAGCAAGGACC	1633
Db	3047	AGAACCGGAGATCCTGCGCGAGCCGCTGACCGCGGTGTACTACGACCCCGCAGCAAGGACC	3106
QY	1634	TGTTGGCCGAGATCCAGAAAGCAGGGCCACGACCAAGTGGACCTTACAGATCTTACCAAGGAC	1693
Db	3107	TGTTGGCCGAGATCCAGAAAGCAGGGCCACGACCAAGTGGACCTTACAGATCTTACCAAGGAC	3166
QY	1694	CCTTCAAGAACCTGAAAGACCGGCAAGTACGCGCAAGTGGCAGCGCACCGCCACACCAACGAG	1753
Db	3167	CCTTCAAGAACCTGAAAGACCGGCAAGTACGCGCAAGTGGCAGCGCACCGCCACACCAACGAG	3226
QY	1754	TGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGGAGAGCATCTGTATCTGGGGCA	1813
Db	3227	TGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGGAGAGCATCTGTATCTGGGGCA	3286
QY	1814	AGACCCCAAGTTTCGCTGCTCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACT	1873
Db	3287	AGACCCCAAGTTTCGCTGCTCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACT	3346
QY	1874	ACTGGCAGGCACCTGTGATCCCGAGTGGGAGTGTGTGAACACCCCCCGCTGGTGAAGC	1933
Db	3347	ACTGGCAGGCACCTGTGATCCCGAGTGGGAGTGTGTGAACACCCCCCGCTGGTGAAGC	3406
QY	1934	TGTGTACACAGCTGGAGAAAGGAGCCATCATCGGCGCGAGACCTTCTACCTGGACCGCG	1993
Db	3407	TGTGTACACAGCTGGAGAAAGGAGCCATCATCGGCGCGAGACCTTCTACCTGGACCGCG	3466
QY	1994	CCGCCAACCGCGAGACCAAGATTCGGCAAGGCGCGCTACGTGACCGACCGGGCGCGGCA	2053
Db	3467	CCGCCAACCGCGAGACCAAGATTCGGCAAGGCGCGCTACGTGACCGACCGGGCGCGGCA	3526
QY	2054	AGATCGTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATTCAGCTGG	2113
Db	3527	AGATCGTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATTCAGCTGG	3586
QY	2114	CCCTGCAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCGACAGCGCAGTACGCCCTGGGCA	2173
Db	3587	CCCTGCAGGACAGCGGCGAGCGAGGTGAACATCTGTGACCGACAGCGCAGTACGCCCTGGGCA	3646
QY	2174	TCATCCAGGCCAGCCCGCAAGAGCGAGAGCGAGCTGGTGAACACAGATCATTCGAGCAGC	2233
Db	3647	TCATCCAGGCCAGCCCGCAAGAGCGAGAGCGAGCTGGTGAACACAGATCATTCGAGCAGC	3706
QY	2234	TGATCAAGAGAGAGAGTGTACTTGTAGCTGGGTGCCCGCCCAACAGGGGCATTCGCGGCA	2293
Db	3707	TGATCAAGAGAGAGAGTGTACTTGTAGCTGGGTGCCCGCCCAACAGGGGCATTCGCGGCA	3766
QY	2294	ACGAGCAGATCGACAAAGCTGGTGAGCAAGGGCATCCCGCAAGGTGCTGTTCTCTGGAGCGCA	2353

Db 3767 ACAGCAGATCGACAAGCTGTGTGACGAAGGGCATCCGCAAGGTGCTGTTCTTGGACGGCA 3826  
Qy 2354 TCGATGGCGGCATCGTGTATCTTACCAAGTACATGAGACCTGTAGTGGGAGCGCGGCC 2413  
Db 3827 TCGATGGCGGCATCGTGTATCTTACCAAGTACATGAGACCTGTATCGTGGGAGCGCGGCC 3886  
Qy 2414 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 3887 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3924

## RESULT 5

US-09-899-575-31  
; Sequence 31, Application US/09899575  
; Publication No. US20030223961A1  
; GENERAL INFORMATION:  
; APPLICANT: Zur Megede, Jan  
; APPLICANT: Barnett, Susan W.  
; APPLICANT: Egnelbrecht, Susan  
; APPLICANT: van Rensburg, Estrelita Janse  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF  
; FILE REFERENCE: PP01631.102  
; CURRENT APPLICATION NUMBER: US/09/899,575  
; CURRENT FILING DATE: 2001-07-05  
; PRIOR APPLICATION NUMBER: 09/475,704  
; PRIOR FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 135  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 31  
; LENGTH: 2463  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PR975YM  
US-09-899-575-31

Query Match 99.2%; Score 2436.2; DB 10; Length 2463;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
Qy 1 GTCAGCCACCATGGCGAGGCCATGAGCGAGCCACCGAGCGCCACCATCTGTATGACG 60  
Db 1 GTCAGCCACCATGGCGAGGCCATGAGCGAGCCACCGAGCGCCACCATCTGTATGACG 60  
Qy 61 GCGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGGGCAAGAGGGC 120  
Db 61 GCGAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGGGCAAGAGGGC 120  
Qy 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
Db 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180  
Qy 181 GGCCACCAAGATGAAGGACTGACCGAGCGCGCGAGCCCAACTTCTTCCGCGAGGACCTGGCC 240  
Db 181 GGCCACCAAGATGAAGGACTGACCGAGCGCGCGAGCCCAACTTCTTCCGCGAGGACCTGGCC 240  
Qy 241 TTCCCCCAAGGAGGCGCGGAGTTCCCAAGCGAGGAGAACCGCGCCCAACAGCCCCCACC 300  
Db 241 TTCCCCCAAGGAGGCGCGGAGTTCCCAAGCGAGGAGAACCGCGCCCAACAGCCCCCACC 300  
Qy 301 AGCCGCGAGCTGAGGTGGCGGAGCAACCCCGAGCGAGGCGCGGCGCGAGCGCCAG 360  
Db 301 AGCCGCGAGCTGAGGTGGCGGAGCAACCCCGAGCGAGGCGCGGCGCGAGCGCCAG 360  
Qy 361 GGCAACCTGAACTTCCCCAGATCACTGTGGCAGCGCCCGCTGTGTGAGCATCAAGGTG 420  
Db 361 GGCAACCTGAACTTCCCCAGATCACTGTGGCAGCGCCCGCTGTGTGAGCATCAAGGTG 420  
Qy 421 GGCGGCGAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGCACCGTGTCTGGAGGAG 480  
Db 421 GGCGGCGAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGCACCGTGTCTGGAGGAG 480

Qy 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAG 540  
Db 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTTCATCAAG 540  
Qy 541 GTGGCGCATGACGACGAGATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
Db 541 GTGGCGCATGACGACGAGATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAACCGTG 600  
Qy 601 CTGATCGGCGCCACCCCGGTGAACATCATCGGCGCAACATGCTGAGCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCGCCACCCCGGTGAACATCATCGGCGCAACATGCTGAGCCAGCTGGGCTGC 660  
Qy 661 ACCCTGAACCTTCCCGCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720  
Db 661 ACCCTGAACCTTCCCGCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720  
Qy 721 GACGCGCCCAAGGTGAAGAGAGTGGCCCTTGACCGAGAGAGATCAAGGCGCTTACCGGCC 780  
Db 721 GACGCGCCCAAGGTGAAGAGAGTGGCCCTTGACCGAGAGAGATCAAGGCGCTTACCGGCC 780  
Qy 781 ATCTGCGAGGAGATGGAGAGAGGAGGCAAGATCAACCAAGATCGGCGCCCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGATGGAGAGAGGAGGCAAGATCAACCAAGATCGGCGCCCGAGAACCCCTAC 840  
Qy 841 AACACCCCGCTGTTCGCCATCAAGAAGAAGACAGCAGCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGCTGTTCGCCATCAAGAAGAAGACAGCAGCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCCGCGAGTGAACAAGCGCACCCAGGACTTTCGGAGGTGACAGTGGGCGATCCCCCAC 960  
Db 901 TTCCGCGAGTGAACAAGCGCACCCAGGACTTTCGGAGGTGACAGTGGGCGATCCCCCAC 960  
Qy 961 CCGCGCGCTGGAAGAGAGAGAGCGTGACCGTGTGGACGTGGGCGAGCGCTACTTTC 1020  
Db 961 CCGCGCGCTGGAAGAGAGAGAGCGTGACCGTGTGGACGTGGGCGAGCGCTACTTTC 1020  
Qy 1021 AGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCCCAGCATCAAC 1080  
Db 1021 AGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTTCACATCCCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCGCCAGGCTGGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGGCATCCGCTACCAAGTACAAAGTGTGCGCCAGGCTGGAAGGGCAGC 1140  
Qy 1141 CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCCC 1200  
Db 1141 CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCCC 1200  
Qy 1201 GAGATCGTGTATCTACCAAGCGCCCTCTGTAGTGGGCGAGCGACCTGGAGATCGGCGCAGC 1260  
Db 1201 GAGATCGTGTATCTACCAAGCGCCCTCTGTAGTGGGCGAGCGACCTGGAGATCGGCGCAGC 1260  
Qy 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCAGCATGCTGCTGCGCTGGGCTTTCACACCCCGAC 1320  
Db 1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCAGCATGCTGCTGCGCTGGGCTTTCACACCCCGAC 1320  
Qy 1321 AAGAGCACAGAGAGGAGCGCCCTTCTGCGCCAT-----CGAGCTGCACCCCGACAG 1374  
Db 1321 AAGAGCACAGAGAGGAGCGCCCTTCTGCGCCAT-----CGAGCTGCACCCCGACAG 1374  
Qy 1375 TGGACCGTGCAGCGCCATCGAGCTGCGCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1434  
Db 1375 TGGACCGTGCAGCGCCATCGAGCTGCGCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1434  
Qy 1435 AAGCTGTGGGCGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCAG 1494  
Db 1435 AAGCTGTGGGCGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCAG 1494  
Qy 1495 CTGTGCAAGTGTGCGCGCGCCCAAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1554  
Db 1495 CTGTGCAAGTGTGCGCGCGCCCAAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1554  
Qy 1501 CTGTGCAAGTGTGCGCGCGCCCAAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1560  
Db 1501 CTGTGCAAGTGTGCGCGCGCCCAAGGCGCTGACCGACATCTGTGCGCCCTGACCGAGGAG 1560







```
QY 794 TGGAGAGGAGGCGAAGATCAACCAAGATCGGCGCCCGGAGAACCCCTTAAACACCCCGGTGT 853
Db 2267 TGGAGAGGAGGCGAAGATCAACCAAGATCGGCGCCCGGAGAACCCCTTAAACACCCCGGTGT 2326
QY 854 TCGCCATCAAGAAGAGGAGAGCAAGCAAGTGGCGCGCAAGCTGCTGAGCTTCCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAAGAGGAGAGCAAGCAAGTGGCGCGCAAGCTGCTGAGCTTCCGCGAGCTGA 2386
QY 914 ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGCTGGGCATCCCGACCCCGCGCGCTGA 973
Db 2387 ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGCTGGGCATCCCGACCCCGCGCGCTGA 2446
QY 974 AGAAGAGAGAGAGCGTGACCTGTCTGGAGGTGGGAGCGCTACTTCAAGCTGCGCCCTGG 1033
Db 2447 AGAAGAGAGAGAGCGTGACCTGTCTGGAGGTGGGAGCGCTACTTCAAGCTGCGCCCTGG 2506
QY 1034 ACAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCG 1093
Db 2507 ACAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAACAAAGAGACCCCG 2566
QY 1094 GCATCCGCTACCAAGTACAAAGTGTCTGCCCCAGGGCTGGAGGGAGCGCCCAAGCATCTTCC 1153
Db 2567 GCATCCGCTACCAAGTACAAAGTGTCTGCCCCAGGGCTGGAGGGAGCGCCCAAGCATCTTCC 2626
QY 1154 AGAGCAGCATGACCAAGATCTTGGAGCTTCCGCGCCCGCAACCCCGAGATCGTATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCTTGGAGCTTCCGCGCCCGCAACCCCGAGATCGTATCT 2686
QY 1214 ACCAGGCCCCCTGTACGTGGGAGCGGACCTGGAGATCGGCGACGACCGCGCCCAAGATCG 1273
Db 2687 ACCAGGCCCCCTGTACGTGGGAGCGGACCTGGAGATCGGCGACGACCGCGCCCAAGATCG 2746
QY 1274 AGGAGCTGCGCAAGCACCTGTCTGGGCTGGGGCTTTCACCACCCCGCAAGAGCACCAAGA 1333
Db 2747 AGGAGCTGCGCAAGCACCTGTCTGGGCTGGGGCTTTCACCACCCCGCAAGAGCACCAAGA 2806
QY 1334 AGGAGCCCCCTTCTGCCCCATCGAGCTGCAACCCCGCAAGTGAACCGTGAGCGCCATCG 1393
Db 2807 AGGAGCCCCCTTCTGCCCCATCGAGCTGCAACCCCGCAAGTGAACCGTGAGCGCCATCG 2866
QY 1394 AGCTGCCCGAGAGGAGCTGGACCTGTGAACGACATCCAGAACCTGGTGGGCAAGCTGA 1453
Db 2867 AGCTGCCCGAGAGGAGCTGGACCTGTGAACGACATCCAGAACCTGGTGGGCAAGCTGA 2926
QY 1454 ACTGGGCGAGCCAGATCTACCCCGCATCAAGTGGCGGCTGTGAAGCTGTGGCGG 1513
Db 2927 ACTGGGCGAGCCAGATCTACCCCGCATCAAGTGGCGGCTGTGAAGCTGTGGCGG 2986
QY 1514 GCGCAAGGCGCTGACCGACATCGTGGCCCTGACCGAGGCGGAGCTGGAGCTGGCGG 1573
Db 2987 GCGCAAGGCGCTGACCGACATCGTGGCCCTGACCGAGGCGGAGCTGGAGCTGGCGG 3046
QY 1574 AGAACCGGAGATCTGCGGAGCCCGTGACCGGCTGTACTAGACCCCGCAAGGAGACC 1633
Db 3047 AGAACCGGAGATCTGCGGAGCCCGTGACCGGCTGTACTAGACCCCGCAAGGAGACC 3106
QY 1634 TGGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTTACAGATCTTACAGAGC 1693
Db 3107 TGGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGACCTTACAGATCTTACAGAGC 3166
QY 1694 CTTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCGACCGCCGACACCAACGAGC 1753
Db 3167 CTTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCGACCGCCGACACCAACGAGC 3226
QY 1754 TGAAGCAGCTGACCGGCGGTGAGAGATCGCCATGGAGAGCATGTGATCTGGGGCA 1813
Db 3227 TGAAGCAGCTGACCGGCGGTGAGAGATCGCCATGGAGAGCATGTGATCTGGGGCA 3286
QY 1814 AGACCCCAAGTCTCGGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 1873
Db 3287 AGACCCCAAGTCTCGGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACT 3346
QY 1874 ACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCTGTGAGC 1933
```

```
Db 3347 ACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACACCCCGCTGTGAGAC 3406
QY 1934 TGTGGTACCAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGAGACGCG 1993
Db 3407 TGTGGTACCAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGAGACGCG 3466
QY 1994 CCGCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGCGCGGAGCA 2053
Db 3467 CCGCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGCGCGGAGCA 3526
QY 2054 AGATCGTGGCTGACCGAGACCAACCAACCAAGAGACCGAGCTGAGGCCCATTCAGCTGG 2113
Db 3527 AGATCGTGGCTGACCGAGACCAACCAACCAAGAGACCGAGCTGAGGCCCATTCAGCTGG 3586
QY 2114 CCTGTCAGGACAGCGCGGAGCGAGGTGAACATCGTGAACCGAGCAGTACGCCCTTGGGCA 2173
Db 3587 CCTGTCAGGACAGCGCGGAGCGAGGTGAACATCGTGAACCGAGCAGTACGCCCTTGGGCA 3646
QY 2174 TCATCCAGGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCGAGATCATCGAGCAGC 2233
Db 3647 TCATCCAGGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCGAGATCATCGAGCAGC 3706
QY 2234 TGATCAAGAGGAGAAAGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATTCGGCGCA 2293
Db 3707 TGATCAAGAGGAGAAAGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATTCGGCGCA 3766
QY 2294 ACAGAGCAGATCGACAAGCTGGTGAAGAGGCGCATCCGCAAGGTCTGTTCCTGAGACGCA 2353
Db 3767 ACAGAGCAGATCGACAAGCTGGTGAAGAGGCGCATCCGCAAGGTCTGTTCCTGAGACGCA 3826
QY 2354 TCAGTGGCGCATCGTGTATCTACAGTACATGAGACGACCTGTACGTGGGCGAGCGCGGCC 2413
Db 3827 TCAGTGGCGCATCGTGTATCTACAGTACATGAGACGACCTGTACGTGGGCGAGCGCGGCC 3886
QY 2414 CTAGATCGATTTAAAGCTTCCCGGGCTAGCACCGGT 2451
Db 3887 CTAGATCGATTTAAAGCTTCCCGGGCTAGCACCGGT 3924
```

RESULT 7  
US-10-190-435-11  
; Sequence 11, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Estrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 3930  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: GagComplPolmutIna\_C  
US-10-190-435-11

Query Match 99.1%; Score 2434.8; DB 15; Length 3930;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 14 TGCCGAGGCGCATGAGCCAGGCGCCACAGCGCCACATCTGTGATGCGAGCGCAACTTCA 73  
Db 1487 TCGCCGAGGCGCATGAGCCAGGCGCCACAGCGCCACATCTGTGATGCGAGCGCAACTTCA 1546

QY 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGGCAAGAGGCGCCACATCGCCGCA 133  
DB |||||  
1547 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGGGCAAGAGGCGCCACATCGCCGCA 1606  
QY 134 ACTGGCGCCCCCGCAAGAAAGGCTGTCTGAAGTGGCGCAAGAGGCGCCACAGATGA 193  
DB |||||  
1607 ACTGGCGCCCCCGCAAGAAAGGCTGTCTGAAGTGGCGCAAGAGGCGCCACAGATGA 1666  
QY 194 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGGGAGGACCTGGCTTCCGCCAGGGCA 253  
DB |||||  
1667 AGGACTGCACCGAGCGCCAGGCCAACTTCTTCGGGAGGACCTGGCTTCCGCCAGGGCA 1726  
QY 254 AGGCGCGGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCGCCCAACAGCGCGCGAGCTGC 313  
DB |||||  
1727 AGGCGCGGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCGCCCAACAGCGCGCGAGCTGC 1786  
QY 314 AGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
DB |||||  
1787 AGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
QY 374 TCCCCAGATCACTCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGGCGCGAGTCA 433  
DB |||||  
1847 TCCCCAGATCACTCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGGCGCGAGTCA 1906  
QY 434 AGGAGCCCTGTGGACACCGCGCGCGAGCACCGTGTGGAGAGATGAGCCTGCGCG 493  
DB |||||  
1907 AGGAGCCCTGTGGCAGCGCGCGCGAGCACCGTGTGGAGAGATGAGCCTGCGCG 1966  
QY 494 GCAAGTGAAGCCCAAGATGATCGCGGCGCATCGCGGCTTCAATCAAGGTGGCGCGAGTAG 553  
DB |||||  
1967 GCAAGTGAAGCCCAAGATGATCGCGGCGCATCGCGGCTTCAATCAAGGTGGCGCGAGTAG 2026  
QY 554 ACCAGATCTGTATCAGATCTCGGGCAGAGAGCCATCGGCACCGTGTGATCGGCCCA 613  
DB |||||  
2027 ACCAGATCTGTATCAGATCTCGGGCAGAGAGCCATCGGCACCGTGTGATCGGCCCA 2086  
QY 614 CCCCGTGAACATCATCGCGCGCAATGCTGACCCAGCTGGGCTGCACTTCAAGTTC 673  
DB |||||  
2087 CCCCGTGAACATCATCGCGCGCAATGCTGACCCAGCTGGGCTGCACTTCAAGTTC 2146  
QY 674 CCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCATGGAAGCGCCCAAGG 733  
DB |||||  
2147 CCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCATGGAAGCGCCCAAGG 2206  
QY 734 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCATCTCGAGAGGA 793  
DB |||||  
2207 TGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCATCTCGAGAGGA 2266  
QY 794 TGGAGAGGAGGCAAGATCAACAAGATCGGGCCCGAGAACCCCTACAAACCCCGTGT 853  
DB |||||  
2267 TGGAGAGGAGGCAAGATCAACAAGATCGGGCCCGAGAACCCCTACAAACCCCGTGT 2326  
QY 854 TCGCCATCAAGAAGAGGACGACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 913  
DB |||||  
2327 TCGCCATCAAGAAGAGGACGACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGA 2386  
QY 914 ACAAGCGCACCGAGACTTCTGGAGGTGACCTGGGCATCCCCCAACCCCGCGCGCTGA 973  
DB |||||  
2387 ACAAGCGCACCGAGACTTCTGGAGGTGACCTGGGCATCCCCCAACCCCGCGCGCTGA 2446  
QY 974 AGAAGAAGAGCGGTGACCGTGTGGAGTGGGCGAGCGCTTCAAGCGTGGCCCTGG 1033  
DB |||||  
2447 AGAAGAAGAGCGGTGACCGTGTGGAGTGGGCGAGCGCTTCAAGCGTGGCCCTGG 2506  
QY 1034 ACGAGACTTCCGCAAGTACACCGCTTCAACATCCCCAGCATCAACACAGAGACCCCG 1093  
DB |||||  
2507 ACGAGACTTCCGCAAGTACACCGCTTCAACATCCCCAGCATCAACACAGAGACCCCG 2566  
QY 1094 GCATCGCTACAGTACAAAGTGTGCTGCCCCAGGGCTGGAAGGGAGCCCCAGCATCTTCC 1153  
DB |||||  
2567 GCATCGCTACAGTACAAAGTGTGCTGCCCCAGGGCTGGAAGGGAGCCCCAGCATCTTCC 2626  
QY 1154 AGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCT 1213

DB 2627 AGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCT 2686  
QY 1214 ACCAGGCCCCCTGTACTGTGGGCGAGCACTTGGAGATCGGCGAGACACCGCGCAAGATCG 1273  
DB |||||  
2687 ACCAGGCCCCCTGTACTGTGGGCGAGCACTTGGAGATCGGCGAGACACCGCGCAAGATCG 2746  
QY 1274 AGGAGCTGGCAAGCACCTGTGCGTGGGCTTCCACCAACCCCGCAAGAGACCAAGAC 1333  
DB |||||  
2747 AGGAGCTGGCAAGCACCTGTGCGTGGGCTTCCACCAACCCCGCAAGAGACCAAGAC 2806  
QY 1334 AGGAGCCCCCTTCTGCCCCATCGAGCTGCACCCCGCAACAGTGAACCGTGCAGCCCCATCG 1393  
DB |||||  
2807 AGGAGCCCCCTTCTGCCCCATCGAGCTGCACCCCGCAACAGTGAACCGTGCAGCCCCATCG 2866  
QY 1394 AGCTGCCGAGAGAGAGAGCTGGACCGTGAACGACATCAAGAGCTGGTGGGCAAGCTGA 1453  
DB |||||  
2867 AGCTGCCGAGAGAGAGAGCTGGACCGTGAACGACATCAAGAGCTGGTGGGCAAGCTGA 2926  
QY 1454 ACTGGGCGAGCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGAAGCTGCTGCGCG 1513  
DB |||||  
2927 ACTGGGCGAGCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGAAGCTGCTGCGCG 2986  
QY 1514 GCGCAGGCGCTGACCGACATCGTGCCTTGAACCGAGGCGCGAGCTGAGAGCTGGCGG 1573  
DB |||||  
2987 GCGCAGGCGCTGACCGACATCGTGCCTTGAACCGAGGCGCGAGCTGAGAGCTGGCGG 3046  
QY 1574 AGAACCGGAGATCTCGCGAGCGCGTGCACGGGTGTACTGAACCCAGCAAGGAGCC 1633  
DB |||||  
3047 AGAACCGGAGATCTCGCGAGCGCGTGCACGGGTGTACTGAACCCAGCAAGGAGCC 3106  
QY 1634 TGGTGGCGAGATCCAGAGCAGGCGCACGACCTGAGTGAACCTTACAGATCTTACAGGAGC 1693  
DB |||||  
3107 TGGTGGCGAGATCCAGAGCAGGCGCACGACCTGAGTGAACCTTACAGATCTTACAGGAGC 3166  
QY 1694 CTTTCAAGAACCTGAAAGCCGCGCAAGTACGCCAAGATCGCACCCGCCACACCAACGAGC 1753  
DB |||||  
3167 CTTTCAAGAACCTGAAAGCCGCGCAAGTACGCCAAGATCGCACCCGCCACCAACGAGC 3226  
QY 1754 TGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATCTGTGATCTGGGGA 1813  
DB |||||  
3227 TGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCATCTGTGATCTGGGGA 3286  
QY 1814 AGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTTGTGGAGCCGACT 1873  
DB |||||  
3287 AGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTTGGAGAGACCTTGTGGAGCCGACT 3346  
QY 1874 ACTGGCAGGCGCACCTTGAATCCCGAGTGGGAGTTCGTGAACAACCCCGCTGTGTAAGC 1933  
DB |||||  
3347 ACTGGCAGGCGCACCTTGAATCCCGAGTGGGAGTTCGTGAACAACCCCGCTGTGTAAGC 3406  
QY 1934 TGTGTGTAACAGTGTGAAGAGGAGCCATCATCGCGCGCGAGACCTTCTACGTGGAAGCGG 1993  
DB |||||  
3407 TGTGTGTAACAGTGTGAAGAGGAGCCATCATCGCGCGCGAGACCTTCTACGTGGAAGCGG 3466  
QY 1994 CCGCCAAACCGGAGACCAAGATCCGCAAGGCGGCTAGTGACCGGCGGCGGCGGCGAG 2053  
DB |||||  
3467 CCGCCAAACCGGAGACCAAGATCCGCAAGGCGGCTAGTGACCGGCGGCGGCGGCGAG 3526  
QY 2054 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGG 2113  
DB |||||  
3527 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGG 3586  
QY 2114 CCTTCAAGGACAGCGGCGAGGAGTGAACATCTGTGAACGACAGCCAGCTACGCCCTGGGA 2173  
DB |||||  
3587 CCTTCAAGGACAGCGGCGAGGAGTGAACATCTGTGAACGACAGCCAGCTACGCCCTGGGA 3646  
QY 2174 TCATCCAGGCGGCGGCGAGGAGGAGCTGTGAACAGATCATCGAGCAGC 2233  
DB |||||  
3647 TCATCCAGGCGGCGGCGGAGGAGGAGCTGTGAACAGATCATCGAGCAGC 3706  
QY 2234 TGATCAAGAGGAGAGGTGTACTTGAAGTGGGTGCGCGCCGCCACCAAGGCGCATCGCGGCA 2293

Db 3707 TGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAACAAGGGCATCGCGGGCA 3766  
Qy 2294 ACGAGCAGATCGACAAGCTGGTGAAGCAAGGCATCCGCAAGGTGCTTCTTGGAGCGCA 2353  
Db 3767 ACGAGCAGATCGACAAGCTGGTGAAGCAAGGCATCCGCAAGGTGCTTCTTGGAGCGCA 3826  
Qy 2354 TCGATGGCGGCATCGTGTATCTACCAAGTACATGAGGACCTGTAGTGGGAGCGCGGCC 2413  
Db 3827 TCGATGGCGGCATCGTGTATCTACCAAGTACATGAGGACCTGTAGTGGGAGCGCGGCC 3886  
Qy 2414 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db 3887 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3924

RESULT 8  
US-10-190-435-58  
; Sequence 58: Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBURG, Esrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: P18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 58  
; LENGTH: 5184  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: TatRevNefagCpolIna C  
US-10-190-435-58

Query Match 99.1%; Score 2434.8; DB 15; Length 5184;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCGCATGAGCGAGCCACCGAGCCATCTGATCAGCGAGCAACTTCA 73  
Db 2741 TCGCGAGGCGCATGAGCGAGCCACCGAGCCATCTGATGAGCGAGCAACTTCA 2800  
Qy 74 AGGCGCCCAAGCGCATCATCAAGTGTTCATCTCGGCAAGGAGGCGCATCGCCGCA 133  
Db 2801 AGGCGCCCAAGCGCATCATCAAGTGTTCATCTCGGCAAGGAGGCGCATCGCCGCA 2860  
Qy 134 ACTGCGCGCGCCCGGAGAGGCGCTGCGAGTGGGAGGCGGAGGCGGCGGCGGAGTGA 193  
Db 2861 ACTGCGCGCGCCCGGAGAGGCGCTGCGAGTGGGAGGCGGAGGCGGCGGAGTGA 2920  
Qy 194 AGGACTGACCGAGCGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCG 253  
Db 2921 AGGACTGACCGAGCGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCG 2980  
Qy 254 AGGCGCGGAGTTCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 313  
Db 2981 AGGCGCGGAGTTCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3040  
Qy 314 AGGTCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 373  
Db 3041 AGGTCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3100  
Qy 374 TCCCGCCAGATCACCTGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 433  
Db 3101 TCCCGCCAGATCACCTGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3160  
Qy 434 AGGAGGCGCTGCTGGAGACCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 493

Db 3161 AGGAGGCGCTGCTGGGCGCACCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3220  
Qy 494 GCAAGTGGAAAGCCCAAGATGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 553  
Db 3221 GCAAGTGGAAAGCCCAAGATGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3280  
Qy 554 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 613  
Db 3281 ACCAGATCTCTGATCGAGATCTGCGGCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3340  
Qy 614 CCGCGGTGAACATCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 673  
Db 3341 CCGCGGTGAACATCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3400  
Qy 674 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCGGCGGCGGCGGCGGCGGCGGCG 733  
Db 3401 CCATCAGCCCCCATCGAGACCGTGCCTGGAAGCTGGAAGCGGCGGCGGCGGCGGCGGCGG 3460  
Qy 734 TGAAGCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 793  
Db 3461 TGAAGCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3520  
Qy 794 TGAAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 853  
Db 3521 TGAAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3580  
Qy 854 TCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 913  
Db 3581 TCGCCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3640  
Qy 914 ACAAGCGCACCCAGGAGCTTCTGGGAGGTGCGAGCTGGGCGATCCCCACCGCGGCGGCGG 973  
Db 3641 ACAAGCGCACCCAGGAGCTTCTGGGAGGTGCGAGCTGGGCGATCCCCACCGCGGCGGCG 3700  
Qy 974 AGAAGAGAGAGGCGTGAACCGTGTGAGCGTGGGCGAGCGGCTACTTCAAGCGTGGCGGCG 1033  
Db 3701 AGAAGAGAGAGGCGTGAACCGTGTGAGCGTGGGCGAGCGGCTACTTCAAGCGTGGCGG 3760  
Qy 1034 ACCAGGAGCTTCCGCAAGTACACCGCTTCCACCATCCCGGCGGCGGCGGCGGCGGCGGCG 1093  
Db 3761 ACCAGGAGCTTCCGCAAGTACACCGCTTCCACCATCCCGGCGGCGGCGGCGGCGGCGG 3820  
Qy 1094 GCATCCGCTACCAAGTACCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1153  
Db 3821 GCATCCGCTACCAAGTACCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3880  
Qy 1154 AGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 1213  
Db 3881 AGAGCAGCATGACCAAGATCTTGGAGCGCTTCCGCGCGGCGGCGGCGGCGGCGGCGGCGG 3940  
Qy 1214 ACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1273  
Db 3941 ACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 4000  
Qy 1274 AGGAGCTGCGCAAGACACCTGCTGCGTGGGCGGCTTCCACCGCGGCGGCGGCGGCGGCG 1333  
Db 4001 AGGAGCTGCGCAAGACACCTGCTGCGTGGGCGGCTTCCACCGCGGCGGCGGCGGCGGCG 4060  
Qy 1334 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1393  
Db 4061 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4120  
Qy 1394 AGCTGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1453  
Db 4121 AGCTGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4180  
Qy 1454 ACTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1513  
Db 4181 ACTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4240  
Qy 1514 GCGCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1573

Db 4241 GCGCAAGGCCCTGACGACATCGTGCCCTTCAACGAGGAGCCGAGCTGGAGCTGGCGG 4300  
Qy 1574 AGAACCGGAGATCTTGGCGGAGCCCGTGCACGGGTGTACTACGACCCGACGAGGACC 1633  
Db 4301 AGAACCGGAGATCTTGGCGGAGCCCGTGCACGGGTGTACTACGACCCGACGAGGACC 4360  
Qy 1634 TGGTGGCCGAGATCCAGAAGCAGGSCCAGACAGTGGACCTTACCAGATCTTACCAGGAGC 1693  
Db 4361 TGGTGGCCGAGATCCAGAAGCAGGSCCAGACAGTGGACCTTACCAGATCTTACCAGGAGC 4420  
Qy 1694 CTTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATGCGCACCCGCCCAACACGAGC 1753  
Db 4421 CTTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATGCGCACCCGCCCAACACGAGC 4480  
Qy 1754 TGAACGAGTGAACGAGGCGGTGCGAGAGATGCCATGAGAGGATCTGTGATCTGGGGCA 1813  
Db 4481 TGAACGAGTGAACGAGGCGGTGCGAGAGATGCCATGAGAGGATCTGTGATCTGGGGCA 4540  
Qy 1814 AGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACT 1873  
Db 4541 AGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACT 4600  
Qy 1874 ACTGCAGGCCACCTGGATCCCGGAGTGGAGTTTCGTGAACACCCCGCCCTGGTGAAGC 1933  
Db 4601 ACTGCAGGCCACCTGGATCCCGGAGTGGAGTTTCGTGAACACCCCGCCCTGGTGAAGC 4660  
Qy 1934 TGTGTACCACTGAGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGGAGCGG 1993  
Db 4661 TGTGTACCACTGAGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGGAGCGG 4720  
Qy 1994 CGGCAACCGCGAGACCAAGATCGGCAAGGCGCGGTACGTGACGACCGGGCGGGCAGA 2053  
Db 4721 CGGCAACCGCGAGACCAAGATCGGCAAGGCGCGGTACGTGACGACCGGGCGGGCAGA 4780  
Qy 2054 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG 2113  
Db 4781 AGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGG 4840  
Qy 2114 CCTCGAGACAGCGGCGAGGCTGAACATCGTGACCGACAGCAGTACGCCCTGGGCA 2173  
Db 4841 CCTCGAGACAGCGGCGAGGCTGAACATCGTGACCGACAGCAGTACGCCCTGGGCA 4900  
Qy 2174 TCATCCAGSCCAGCCGACAGAGCGAGAGGAGCTGTGTAACCAAGATCATCGAGCAGC 2233  
Db 4901 TCATCCAGSCCAGCCGACAGAGCGAGAGGAGCTGTGTAACCAAGATCATCGAGCAGC 4960  
Qy 2234 TGATCAAGAGGAGAGGTGTACTTACCTGGGTGCCGCCCAAGGGCATTCGCGGCA 2293  
Db 4961 TGATCAAGAGGAGAGGTGTACTTACCTGGGTGCCGCCCAAGGGCATTCGCGGCA 5020  
Qy 2294 ACGAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTTCTTGGAGCGCA 2353  
Db 5021 ACGAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTTCTTGGAGCGCA 5080  
Qy 2354 TCGATGGCGGCATCGTGTACTACAGTACATGAGACGACTGTGAGTGGGCGAGCGCGGC 2413  
Db 5081 TCGATGGCGGCATCGTGTACTACAGTACATGAGACGACTGTGAGTGGGCGAGCGCGGC 5140  
Qy 2414 CTAGGATCGATTAAGGCTTCCGGGGCTAGCACCGGT 2451  
Db 5141 CTAGGATCGATTAAGGCTTCCGGGGCTAGCACCGGT 5178

RESULT 9

US-10-190-305A-82  
; Sequence 82, Application US/10190305A  
; Publication No. US20030198621A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: BARNETT, Susan  
; APPLICANT: LIAN, Ying  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE B AND/OR  
; TITLE OF INVENTION: TYPE C POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF

; FILE REFERENCE: 2302-18702 / 18702.002  
; CURRENT APPLICATION NUMBER: US/10/190.305A  
; CURRENT FILING DATE: 2002-07-05  
; NUMBER OF SEQ ID NOS: 93  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 82  
; LENGTH: 5184  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: TatRevNefgagCpolIna C  
US-10-190-305A-82

Query Match 99.1%; Score 2434.8; DB 16; Length 5184;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCCATGAGCCAGGCCACCGCGCAACATCTGTATGCGAGCGCAGCAACTTCA 73  
Db 2741 TCGCCGAGGCCATGAGCCAGGCCACCGCGCAACATCTGTATGCGAGCGCAGCAACTTCA 2800  
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAAATGCGGCAAGGAGGGCCACATTCGCCCGCA 133  
Db 2801 AGGGCCCCAAGCGCATCATCAAGTGTCTTCAAATGCGGCAAGGAGGGCCACATTCGCCCGCA 2860  
Qy 134 ACTGCGCGCGCCCGCGCAAGAGGGCTGCTGGAAGTGGGCAAGGAGGGCCACCGAGTGA 193  
Db 2861 ACTGCGCGCGCCCGCGCAAGAGGGCTGCTGGAAGTGGGCAAGGAGGGCCACCGAGTGA 2920  
Qy 194 AGGACTGCAACGAGCGCGCCAGGCCCAACTTCTTCCGAGGAGCTGGCCCTTCCCGCAGGGCA 253  
Db 2921 AGGACTGCAACGAGCGCGCCAGGCCCAACTTCTTCCGAGGAGCTGGCCCTTCCCGCAGGGCA 2980  
Qy 254 AGGGCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCCACAGCCGCGAGCTGC 313  
Db 2981 AGGGCCGCGAGTTCCTCCAGCGAGCAGAAACCGCGCAACAGCCCCACAGCCGCGAGCTGC 3040  
Qy 314 AGGTGCGCGGCGCAACACCCCGCAGAGGCGGGCGCGGAGGGCCAGGGCACCTTGAATC 373  
Db 3041 AGGTGCGCGGCGCAACACCCCGCAGAGGCGGGCGCGGAGGGCCAGGGCACCTTGAATC 3100  
Qy 374 TCCCCACAGATCACTCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGCGCCAGATCA 433  
Db 3101 TCCCCACAGATCACTCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGCGCCAGATCA 3160  
Qy 434 AGGAGGCCCTTCTGGACACCGCGCGCGACGACACCGTGTCTGGAGAGATGAGCTTGCCTG 493  
Db 3161 AGGAGGCCCTTCTGGACACCGCGCGCGACGACACCGTGTCTGGAGAGATGAGCTTGCCTG 3220  
Qy 494 GCAAGTGAAGGCCCAAGATGATCGGCGCATCGGGGCTTTCATCAAGGTGGCCAGTACG 553  
Db 3221 GCAAGTGAAGGCCCAAGATGATCGGCGCATCGGGGCTTTCATCAAGGTGGCCAGTACG 3280  
Qy 554 ACCAGATCTCTGATCGAGATCTGCGCAAGAGGCCATCGGCAACCGTGTGTGTCGCGCCCA 613  
Db 3281 ACCAGATCTCTGATCGAGATCTGCGCAAGAGGCCATCGGCAACCGTGTGTGTCGCGCCCA 3340  
Qy 614 CCCCCGTGAACATCATTCGCGCGCAACATGTCTGAACCCAGCTGGGTGCAACCTTGAATCTCC 673  
Db 3341 CCCCCGTGAACATCATTCGCGCGCAACATGTCTGAACCCAGCTGGGTGCAACCTTGAATCTCC 3400  
Qy 674 CCATCAGCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGG 733  
Db 3401 CCATCAGCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGGATGAGCGGCCCAAGG 3460  
Qy 734 TGAAGCAGTGGCCCTTGAACCGGAGAGAGATCAAGGCCCTTGAACCGGATGTCGCGAGGAGA 793  
Db 3461 TGAAGCAGTGGCCCTTGAACCGGAGAGAGATCAAGGCCCTTGAACCGGATGTCGCGAGGAGA 3520  
Qy 794 TGGAGAGGAGGGCAAGATCAACCAAGATCGCGCCCGGAGAACCCCTTACAACACCCCGGTGT 853  
Db 3521 TGGAGAGGAGGGCAAGATCAACCAAGATCGCGCCCGGAGAACCCCTTACAACACCCCGGTGT 3580



QY 121 CACATCGCCGCAACTGCGCGCCGCCCCCGCAGAGAGGCTGCTGAAGTGGCGAAGGAG 180  
 Db 121 CACATCGCCGCAACTGCGCGCCGCCCCCGCAGAGAGGCTGCTGAAGTGGCGAAGGAG 180  
 QY 181 GGCACACAGATGAAGGACTGACACAGAGGCGCCAGGCAAACTTCTCCGCGAGGACCTGGCC 240  
 Db 181 GGCACACAGATGAAGGACTGACACAGAGGCGCCAGGCAAACTTCTCCGCGAGGACCTGGCC 240  
 QY 241 TTCCCCCAAGGCAAGGCGCGAGTTCCCAAGCAGAGCAAAACCGCGCCCAACAGCCCAACC 300  
 Db 241 TTCCCCCAAGGCAAGGCGCGAGTTCCCAAGCAGAGCAAAACCGCGCCCAACAGCCCAACC 300  
 QY 301 AGCCGCGAGCTGACAGTGGCGGCGACAAACCCCGCAGGAGGCGCGCGAGCGCCAG 360  
 Db 301 AGCCGCGAGCTGACAGTGGCGGCGACAAACCCCGCAGGAGGCGCGCGAGCGCCAG 360  
 QY 361 GGCACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGGATCAAGGTG 420  
 Db 361 GGCACCTGAACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGGATCAAGGTG 420  
 QY 421 GGCGCCAGATCAAGAGGCGCTGTGGACACCGCGCGCGACGACACCGTGTGGAGGAG 480  
 Db 421 GGCGCCAGATCAAGAGGCGCTGTGGACACCGCGCGCGACGACACCGTGTGGAGGAG 480  
 QY 481 ATGAGCTGCGCGCAAGTGGAGCCCAAGATGATCGGGCGGATCGGGCGCTTCAACAAG 540  
 Db 481 ATGAGCTGCGCGCAAGTGGAGCCCAAGATGATCGGGCGGATCGGGCGCTTCAACAAG 540  
 QY 541 GTGCGCCAGTACAGCAGATCTGATCGAGATCTCGGCAAGAGGCGCATCGGCAACCGTG 600  
 Db 541 GTGCGCCAGTACAGCAGATCTGATCGAGATCTCGGCAAGAGGCGCATCGGCAACCGTG 600  
 QY 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 Db 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 QY 661 ACCCTGAATTTCCCATCAGCCCATCGAGCCCTGAGACCGTGCCTGAACTGAGCCCGGATG 720  
 Db 661 ACCCTGAATTTCCCATCAGCCCATCGAGCCCTGAGACCGTGCCTGAACTGAGCCCGGATG 720  
 QY 721 GACGCGCCCAAGGTGAAGCAGTGGCGCTGACCGAGAGGAGATCAAGGCGCTGACCGCC 780  
 Db 721 GACGCGCCCAAGGTGAAGCAGTGGCGCTGACCGAGAGGAGATCAAGGCGCTGACCGCC 780  
 QY 781 ATCTCGAGGAGATGGAGAAGAGGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTAC 840  
 Db 781 ATCTCGAGGAGATGGAGAAGAGGGCAAGATCAACAGATCGGCGCCCGAGAACCCCTAC 840  
 QY 841 AACACCCCGTGTTCGCCATCAAGAGAGGACAGACCAAGTGGCGCAAGCTGGTGAC 900  
 Db 841 AACACCCCGTGTTCGCCATCAAGAGAGGACAGACCAAGTGGCGCAAGCTGGTGAC 900  
 QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGTGGGCAATCCCCAC 960  
 Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGTGGGCAATCCCCAC 960  
 QY 961 CCGCGCGGCTGAAGAAGAGAGCGTGACCGTGTGGACGCTGGCGAGCGCTTACTTC 1020  
 Db 961 CCGCGCGGCTGAAGAAGAGAGCGTGACCGTGTGGACGCTGGCGAGCGCTTACTTC 1020  
 QY 1021 AGCTGCGCTGGACGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080  
 Db 1021 AGCTGCGCTGGACGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080  
 QY 1081 AACGAGACCCCGGATCCGTTACAGTACACGTTGCTGCCCGAGGCTGGAAGGGCAGC 1140  
 Db 1081 AACGAGACCCCGGATCCGTTACAGTACACGTTGCTGCCCGAGGCTGGAAGGGCAGC 1140  
 QY 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCC 1200  
 Db 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCCC 1200

QY 1201 GAGATCGTGATCTACACGGCCCCCTGTACTGCTGGCAGCGACCTGGAGATCGGCCAGCAC 1260  
 Db 1201 GAGATCGTGATCTACACGGCCCCCTGTACTGCTGGCAGCGACCTGGAGATCGGCCAGCAC 1260  
 QY 1261 CGCGCAAGATCGAGGAGCTCGCAAGCACTGTGCTGGCTGGGGTTTCAACACCCCGAC 1320  
 Db 1261 CGCGCAAGATCGAGGAGCTCGCAAGCACTGTGCTGGCTGGGGTTTCAACACCCCGAC 1320  
 QY 1321 AAGAAACACAGAGAGAGGAGCCCCCTTCTGCTGCCAT-----CGAGCTGACCCCGACAG 1374  
 Db 1321 AAGAAACACAGAGAGGAGCCCCCTTCTGCTGGCTGGGGTTTCAACACCCCGACAG 1380  
 QY 1375 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGACCGTGAACGACATCCAG 1434  
 Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGACCGTGAACGACATCCAG 1440  
 QY 1435 AAGCTGTGGGCAAGCTGAACCTGGGCCACGACAGATCTACCCCGGCATCAAGGTGCGCCAG 1494  
 Db 1441 AAGCTGTGGGCAAGCTGAACCTGGGCCACGACAGATCTACCCCGGCATCAAGGTGCGCCAG 1500  
 QY 1495 CTGTGCAAGCTGTGCGCGCGCCAAAGCCCTGACCGACATCTGCTCCCTGACCCAGAG 1554  
 Db 1501 CTGTGCAAGCTGTGCGCGCGCCAAAGCCCTGACCGACATCTGCTCCCTGACCCAGAG 1560  
 QY 1555 GCCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGCGCTGTAC 1614  
 Db 1561 GCCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGCGCTGTAC 1620  
 QY 1615 TACGACCCCAAGGACCTGCTGGGCCGAGATCCAGAAAGCAGGGCCACGACCAAGTGGACC 1674  
 Db 1621 TACGACCCCAAGGACCTGCTGGGCCGAGATCCAGAAAGCAGGGCCACGACCAAGTGGACC 1680  
 QY 1675 TACGAGATCTACAGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGAAGTGGC 1734  
 Db 1681 TACGAGATCTACAGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGAAGTGGC 1740  
 QY 1735 ACCGCGCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGTCCCATGGAG 1794  
 Db 1741 ACCGCGCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGTCCCATGGAG 1800  
 QY 1795 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGG 1854  
 Db 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGG 1860  
 QY 1855 GAGACTGTGTGACCGACTACTGCGAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAC 1914  
 Db 1861 GAGACTGTGTGACCGACTACTGCGAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAC 1920  
 QY 1915 ACCCGCCCTTGGTGAAGCTGTGTACCGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1974  
 Db 1921 ACCCGCCCTTGGTGAAGCTGTGTACCGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1980  
 QY 1975 ACCTTTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2034  
 Db 1981 ACCTTTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040  
 QY 2035 ACCGAGCGGGCGCGCGAGAGATGTGTGACCTGACCGAGACCAACCAAGAGAGCCGAG 2094  
 Db 2041 ACCGAGCGGGCGCGCGAGAGATGTGTGACCTGACCGAGACCAACCAAGAGAGCCGAG 2100  
 QY 2095 CTGCAAGCCATTCAGCTGCGCCCTGCGAGACAGCGCAGCGAGGTCAACATCGTGACCCAG 2154  
 Db 2101 CTGCAAGCCATTCAGCTGCGCCCTGCGAGACAGCGCAGCGAGGTCAACATCGTGACCCAG 2160  
 QY 2155 AGCCAGTACGCGCTGGGCGATCATCAGGCCCGAGCCCGCAAGAGCGAGAGCGAGCTGGTG 2214  
 Db 2161 AGCCAGTACGCGCTGGGCGATCATCAGGCCCGAGCCCGCAAGAGCGAGAGCGAGCTGGTG 2220  
 QY 2215 AACGAGATCTGAGCAGCTGATCAAGAAAGAGAGGTGTACCTGAGCTGGGTGCCCCC 2274  
 Db 2221 AACGAGATCTGAGCAGCTGATCAAGAAAGAGAGGTGTACCTGAGCTGGGTGCCCCC 2280  
 QY 2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGAAGCTGGTGAGCAAGGGCATCCGCAAG 2334







```

1615 TACGACCCAGCAAGGACTGTGGCGGAGATCCAGAGGCGGCGCAAGCCAGTGGACC 1674
1621 TACGACCCAGCAAGGACTGTGGCGGAGATCCAGAGGCGGCGCAAGCCAGTGGACC 1680
1675 TACGACGATCTACGAGGAGCCCTTCAAGAACCTTGAAGAGCGGCAAGTACGCAAGATGGC 1734
1681 TACGACGATCTACGAGGAGCCCTTCAAGAACCTTGAAGAGCGGCAAGTACGCAAGATGGC 1740
1735 ACCGCCCCACCAACGAGCGTGAAGCAGGTGACCGAGGCGGTGCAAGAGATGCCATGAG 1794
1741 ACCGCCCCACCAACGAGCGTGAAGCAGGTGACCGAGGCGGTGCAAGAGATGCCATGAG 1800
1795 AGCATGCTGATCTGGGGGAGAGCCCGCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGG 1854
1801 AGCATGCTGATCTGGGGGAGAGCCCGCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGG 1860
1855 GAGACCTGTGACCGACCTACTTGGCAGGCGCACTGGATCCCCGAGTGGGAGTTCTGTGAAC 1914
1861 GAGACCTGTGACCGACCTACTTGGCAGGCGCACTGGATCCCCGAGTGGGAGTTCTGTGAAC 1920
1915 ACCCCCCCTGTGTAAGCTGTGTTACGAGTGGAGAAAGAGCCCATCATCGGCGCCGAG 1974
1921 ACCCCCCCTGTGTAAGCTGTGTTACGAGTGGAGAAAGAGCCCATCATCGGCGCCGAG 1980
1975 ACCTTCTACGTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACCTG 2034
1981 ACCTTCTACGTGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACCTG 2040
2035 ACCGACCGGGCGGCGGAGAGATCTGAGCCTGACCGAGACCAACCAAGAGACCGAG 2094
2041 ACCGACCGGGCGGCGGAGAGATCTGAGCCTGACCGAGACCAACCAAGAGACCGAG 2100
2095 TTGCGAGGCGATCCAGTGGCCCTGAGGACGAGCGGAGGAGTGAACATGCTGACCGAC 2154
2101 CTGCGAGGCGATCCAGTGGCCCTGAGGACGAGCGGAGGAGTGAACATGCTGACCGAC 2160
2155 AGCGAGTACGCGCTGGGCGATCATCAGCGCCGAGCGCGGAGGAGCGAGCGAGTGGT 2214
2161 AGCGAGTACGCGCTGGGCGATCATCAGCGCCGAGCGCGGAGGAGCGAGCGAGTGGT 2220
2215 AACGAGATCATCGACGAGCTGATCAAGAGAGAGAGTGTACTCTGAGTGGGTGCGCGCC 2274
2221 AACGAGATCATCGACGAGCTGATCAAGAGAGAGAGTGTACTCTGAGTGGGTGCGCGCC 2280
2275 CACAGAGGCGATCGGCGGCAACGAGCAGATCGAACAAGCTGTGAGCAAGGCGATCCGCAAG 2334
2281 CACAGAGGCGATCGGCGGCAACGAGCAGATCGAACAAGCTGTGAGCAAGGCGATCCGCAAG 2340
2335 GTGCTGTTCTTGGAGCGGATCGATGGCGGCGATCGTATCAAGTACAGTACGAGCGACCTG 2394
2341 GTGCTGTTCTTGGAGCGGATCGATGGCGGCGATCGTATCAAGTACAGTACGAGCGACCTG 2400
2395 TACGTGGGCGAGCGGCGGCGCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451
2401 TACGTGGGCGAGCGGCGGCGCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

```

RESULT 12

```

US-09-899-575-30
; Sequence 30, Application US/09899575
; Publication No. US2003023961A1
; GENERAL INFORMATION:
; APPLICANT: Zur Megede, Jan
; APPLICANT: Barnett, Susan W.
; APPLICANT: Egnelbrecht, Susan
; APPLICANT: van Rensburg, Estrelita Janee
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C
; TITLE OF INVENTION: POLYPEPTIDES, POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: PP01631.102
; CURRENT APPLICATION NUMBER: US/09/899,575
; CURRENT FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 09/475,704

```

```

; PRIOR FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 2469
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PR975(+)
US-09-899-575-30

```

```

Query Match      98.3%; Score 2415.4; DB 10; Length 2469;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

```

```

QY 1 GTCGACGCGACATGCGCGAGGCCATGAGCGGCACAGCGCCACAGCGCCAAATCTGTATGAG 60
DB 1 GTCGACGCGACATGCGCGAGGCCATGAGCGGCACAGCGGCACAGCGCCAAATCTGTATGAG 60
QY 61 GCGACGAACTTCAAGGGCCCAAGGGCATCATCAAGTCTTCACTGCGGCAAGGAGGCG 120
DB 61 GCGACGAACTTCAAGGGCCCAAGGGCATCATCAAGTCTTCACTGCGGCAAGGAGGCG 120
QY 121 CACATCGCCCCCAACTGCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180
DB 121 CACATCGCCCCCAACTGCGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180
QY 181 GGCACACAGATGAAGGACTGCAACGAGCGCGCAAGGCGCAACTCTTCCGCGAGGACCTGGCC 240
DB 181 GGCACACAGATGAAGGACTGCAACGAGCGCGCAAGGCGCAACTCTTCCGCGAGGACCTGGCC 240
QY 241 TTCCCGCAGGCGCAAGGGCCCGGAGTTCCCGCAGCGAGAGAACCGCGCCAAACAGCCCCCACC 300
DB 241 TTCCCGCAGGCGCAAGGGCCCGGAGTTCCCGCAGCGAGAGAACCGCGCCAAACAGCCCCCACC 300
QY 301 AGCGCGCAGCTGCGAGGTGCGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360
DB 301 AGCGCGCAGCTGCGAGGTGCGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGAGCGCCAG 360
QY 361 GGCACCTGTAACTTCCCGCAGATCACTCTGTGCGAGCGCGCCCTGCTGTGAGCATCAAGGTG 420
DB 361 GGCACCTGTAACTTCCCGCAGATCACTCTGTGCGAGCGCGCCCTGCTGTGAGCATCAAGGTG 420
QY 421 GCGCGCAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGAGCACACCGTCTGTGAGGAG 480
DB 421 GCGCGCAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGAGCACACCGTCTGTGAGGAG 480
QY 481 ATGAGCTTGCCTGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGGCTTTCATCAAG 540
DB 481 ATGAGCTTGCCTGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGGCTTTCATCAAG 540
QY 541 GTGCGCGCATGACGACAGATCTGTATCGAGATCTGCGGCGCAAGAGGCCATCGGCGACCGTG 600
DB 541 GTGCGCGCATGACGACAGATCTGTATCGAGATCTGCGGCGCAAGAGGCCATCGGCGACCGTG 600
QY 601 CTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 CTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 ACCCTGAACTTCCCGCATCAGCGCCCATCGAGACCGGTGCGCGCGTGAAGCTGAAGCCCGGCGATG 720
DB 661 ACCCTGAACTTCCCGCATCAGCGCCCATCGAGACCGGTGCGCGCGTGAAGCTGAAGCCCGGCGATG 720
QY 721 GACGCGCGCAAGGAGTGAAGCAGTGGCGCCCTGACCGGAGGAGAGATCAAGGCGCTGACCGCC 780
DB 721 GACGCGCGCAAGGAGTGAAGCAGTGGCGCCCTGACCGGAGGAGAGATCAAGGCGCTGACCGCC 780
QY 781 ATCTGCGAGGAGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
DB 781 ATCTGCGAGGAGATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
QY 841 AACACCCCGGTGTTCCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
DB 841 AACACCCCGGTGTTCCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900

```

Db 841 AACACCCCGTGTTCGCCATCAAGAGAGAGACCAAGAGTGGCGCAAGCTGTGTGAC 900  
Qy 901 TTCGCGAGCTGAACAAGCGCACCCAGAGATCTTGGGAGGTGCAGCTGGGCATCCCCCAC 960  
Db 901 TTCGCGAGCTGAACAAGCGCACCCAGAGATCTTGGGAGGTGCAGCTGGGCATCCCCCAC 960  
Qy 961 CCGCGCGCTGAAGAGAGAGAGCTGACCGTGTGAGCTGTGGGCGAGCGCTACTTTC 1020  
Db 961 CCGCGCGCTGAAGAGAGAGAGCTGACCGTGTGAGCTGTGGGCGAGCGCTACTTTC 1020  
Qy 1021 AGCGTGCCTTGGAGAGAGATCTTCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
Db 1021 AGCGTGCCTTGGAGAGAGATCTTCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGCATCCGTTACAGTACAACTGTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGCGCATCCGTTACAGTACAACTGTGCTGCCCGAGGGCTGGAAGGGCAGC 1140  
Qy 1141 CCGAGCATCTTCAGAGCAGCATGACCAAGATCTGTGAGCGCTTCCGCGCGCCGCAACCCC 1200  
Db 1141 CCGAGCATCTTCAGAGCAGCATGACCAAGATCTGTGAGCGCTTCCGCGCGCCGCAACCCC 1200  
Qy 1201 GAGATCGTGATCTACCA-----GSCCCCTGTGTACGTGGGCGAGCGACTGGAGATCGGC 1254  
Db 1201 GAGATCGTGATCTACCA-----GSCCCCTGTGTACGTGGGCGAGCGACTGGAGATCGGC 1254  
Qy 1255 CAGCACCGCGCAAGATCAGAGAGCTGCGCAAGACACTGTGCTGCGCTGGGGCTTCAACACCC 1314  
Db 1255 CAGCACCGCGCAAGATCAGAGAGCTGCGCAAGACACTGTGCTGCGCTGGGGCTTCAACACCC 1314  
Qy 1315 CCGCAAGAAGAGCCAGAGAGAGCGCCCTTCTGTCGCCAT-----CGAGCTGCACCCC 1368  
Db 1315 CCGCAAGAAGAGCCAGAGAGAGCGCCCTTCTGTCGCCAT-----CGAGCTGCACCCC 1368  
Qy 1369 GACAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGCTGCACCGTGAACGAC 1428  
Db 1369 GACAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGCTGCACCGTGAACGAC 1428  
Qy 1429 ATCCAGAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTAACCCCGGCATCAAGGTG 1488  
Db 1429 ATCCAGAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTAACCCCGGCATCAAGGTG 1488  
Qy 1489 CGCAGCTGTGCAAGCTGTCTGGCGCGCCCAAGGCCCTTGACCGACATCTGTGCCCTGACC 1548  
Db 1489 CGCAGCTGTGCAAGCTGTCTGGCGCGCCCAAGGCCCTTGACCGACATCTGTGCCCTGACC 1548  
Qy 1549 GAGGAGCGAGCTGGAGCTGGCGAGACCGCGAGATCTTGGCGGAGCGCCGTGCACGGC 1608  
Db 1549 GAGGAGCGAGCTGGAGCTGGCGAGACCGCGAGATCTTGGCGGAGCGCCGTGCACGGC 1608  
Qy 1609 GTGTACTACGACCCAGCAAGGACTGTGTGGCCGAGATCCAGAGAGCGGGCCAGCACGAC 1668  
Db 1609 GTGTACTACGACCCAGCAAGGACTGTGTGGCCGAGATCCAGAGAGCGGGCCAGCACGAC 1668  
Qy 1669 TGGACCTACAGATCTTACAGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728  
Db 1669 TGGACCTACAGATCTTACAGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728  
Qy 1729 ATGCGCACCGCCACCAACGAGCTGAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1788  
Db 1729 ATGCGCACCGCCACCAACGAGCTGAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1788  
Qy 1741 ATGCGCACCGCCACCAACGAGCTGAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1800  
Db 1741 ATGCGCACCGCCACCAACGAGCTGAGCAGCTGACCGAGCGCGTGCAGAGATCGCC 1800  
Qy 1789 ATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTTCGCGCTGCCCATCCAGAGAG 1848  
Db 1789 ATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTTCGCGCTGCCCATCCAGAGAG 1848  
Qy 1849 ACCTGGGAGACCTGGTGGACCGACTTCTGGCAGGCCACCTGGATCCCGAGTGGGAGTTC 1908  
Db 1849 ACCTGGGAGACCTGGTGGACCGACTTCTGGCAGGCCACCTGGATCCCGAGTGGGAGTTC 1908  
Qy 1909 GTGAACACCCCGCTTGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGC 1968  
Db 1909 GTGAACACCCCGCTTGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGC 1968  
Qy 1921 GTGAACACCCCGCTTGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGC 1980  
Db 1921 GTGAACACCCCGCTTGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGC 1980

Qy 1969 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2028  
Db 1969 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGC 2028  
Qy 2029 TACGTGGACCGACCGCGCGCGCAGAGATCGTGGACCTTCCGAGACCAACCAAGAG 2088  
Db 2029 TACGTGGACCGACCGCGCGCGCAGAGATCGTGGACCTTCCGAGACCAACCAAGAG 2088  
Qy 2041 TACGTGGACCGACCGCGCGCGCAGAGATCGTGGACCTTCCGAGACCAACCAAGAG 2100  
Db 2041 TACGTGGACCGACCGCGCGCGCAGAGATCGTGGACCTTCCGAGACCAACCAAGAG 2100  
Qy 2089 ACCGAGCTGACGGGCATCCAGCTGGCCCTGAGGACAGCGGCGAGGTTGAAATCGTG 2148  
Db 2089 ACCGAGCTGACGGGCATCCAGCTGGCCCTGAGGACAGCGGCGAGGTTGAAATCGTG 2148  
Qy 2101 ACCGAGCTGACGGGCATCCAGCTGGCCCTGAGGACAGCGGCGAGGTTGAAATCGTG 2160  
Db 2101 ACCGAGCTGACGGGCATCCAGCTGGCCCTGAGGACAGCGGCGAGGTTGAAATCGTG 2160  
Qy 2149 ACCGAGCGAGTACGCTTGGGCGCATCATCCAGGCCAGCCCGCAAGAGCGAGCGAG 2208  
Db 2149 ACCGAGCGAGTACGCTTGGGCGCATCATCCAGGCCAGCCCGCAAGAGCGAGCGAG 2208  
Qy 2209 CTGCTGAACCAAGATCATCGAGCAGCTGATCAGAGAGAGAGGTTGACCTGAGCTGGTG 2268  
Db 2209 CTGCTGAACCAAGATCATCGAGCAGCTGATCAGAGAGAGAGGTTGACCTGAGCTGGTG 2268  
Qy 2221 CTGCTGAACCAAGATCATCGAGCAGCTGATCAGAGAGAGAGGTTGACCTGAGCTGGTG 2280  
Db 2221 CTGCTGAACCAAGATCATCGAGCAGCTGATCAGAGAGAGAGGTTGACCTGAGCTGGTG 2280  
Qy 2269 CCGCCCCACAAGGAGTCCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2328  
Db 2269 CCGCCCCACAAGGAGTCCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2328  
Qy 2281 CCGCCCCACAAGGAGTCCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
Db 2281 CCGCCCCACAAGGAGTCCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
Qy 2329 CGCAAGTGTCTTCTTGGAGCGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2388  
Db 2329 CGCAAGTGTCTTCTTGGAGCGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2388  
Qy 2341 CGCAAGTGTCTTCTTGGAGCGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2400  
Db 2341 CGCAAGTGTCTTCTTGGAGCGCATCGATGGCGGCATCGTGATCTACAGTACATGGAC 2400  
Qy 2389 GACCTGTACGTGGGCGAGCGCGCTTGGAGTCAAGATCTTAAAGCTTCCCGGGCTAGCAC 2448  
Db 2389 GACCTGTACGTGGGCGAGCGCGCTTGGAGTCAAGATCTTAAAGCTTCCCGGGCTAGCAC 2448  
Qy 2401 GACCTGTACGTGGGCGAGCGCGCTTGGAGTCAAGATCTTAAAGCTTCCCGGGCTAGCAC 2460  
Db 2401 GACCTGTACGTGGGCGAGCGCGCTTGGAGTCAAGATCTTAAAGCTTCCCGGGCTAGCAC 2460  
Qy 2449 GGTGAATTC 2457  
Db 2449 GGTGAATTC 2457  
Qy 2461 GGTGAATTC 2469  
Db 2461 GGTGAATTC 2469

## RESULT 13

US-10-190-435-13

; Sequence 13, Application US/10190435

; Publication No. US20030143248A1

; GENERAL INFORMATION:

; APPLICANT: ZUR MEDEDE, Jan

; APPLICANT: BARNETT, Susan W.

; APPLICANT: LIAN, Ying

; APPLICANT: ENGELBRECHT, Susan

; APPLICANT: VAN RENSBURG, Estrelita J.

; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C

; FILE REFERENCE: PP18133.003 / 2302-18133

; CURRENT APPLICATION NUMBER: US/10190,435

; CURRENT FILING DATE: 2002-12-30

; NUMBER OF SEQ ID NOS: 319

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 3531

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: GagPolmut\_c

US-10-190-435-13

Query Match

Best Local Similarity 97.98; Score 2404.4; DB 15; Length 3531;

Matches 2417; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 14 TGGCCGAGGCGATGAGCGAGCGCCAGCGCGCAACATCTGTATGCGAGCGAGCAATTC 73

Db 1088 TGGCCGAGGCGATGAGCGAGCGCCAGCGCGCAACATCTGTATGCGAGCGAGCAATTC 1147

Qy 74 AGGCCCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGCGCAATCGCCGCA 133

Db 74 AGGCCCCCAAGCGCATCATCAAGTGTTCATCTGCGGCAAGGAGGCGCAATCGCCGCA 133

Db 1148 AGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGGCGCAAGAGGGCCACATCGCCGCA 1207  
Qy 134 ACTGCGCGCCCCCGCAGAGAGGGCTGCTGAAGTGGCGCAAGAGGGCCACAGATGA 193  
Db 1208 ACTGCGCGCCCCCGCAGAGAGGGCTGCTGAAGTGGCGCAAGAGGGCCACAGATGA 1267  
Qy 194 AGGACTGCACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCCCTTCCCGCCAGGCA 253  
Db 1268 AGGACTGCACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCCCTTCCCGCCAGGCA 1327  
Qy 254 AGGCGCGAGTTCCTCCAGCGAGAGAAACCGCGCAACAGCGCCCAACAGCGCGCGAGCTGC 313  
Db 1328 AGGCGCGAGTTCCTCCAGCGAGAGAAACCGCGCAACAGCGCCCAACAGCGCGCGAGCTGC 1387  
Qy 314 AGGTGCGGCGCAACAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
Db 1398 AGGTGCGGCGCAACAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1447  
Qy 374 TCCCCAGATCACTCTGTGGCAGCGCCCTCTGGTGAGCATCAAGGTGGCGGCGCAGATCA 433  
Db 1448 TCCCCAGATCACTCTGTGGCAGCGCCCTCTGGTGAGCATCAAGGTGGCGGCGCAGATCA 1507  
Qy 434 AGGAGCCTGTGTGACACCGCGCGCAGACACACCGTGTGAGAGATGAGCCTGCCCG 493  
Db 1508 AGGAGCCTGTGTGACACCGCGCGCAGACACACCGTGTGAGAGATGAGCCTGCCCG 1567  
Qy 494 GCAAGTGAAGCCCAAGATGATCGGGCGCATCGGGCTTTCATCAAGGTGGCGCGCATAG 553  
Db 1568 GCAAGTGAAGCCCAAGATGATCGGGCGCATCGGGCTTTCATCAAGGTGGCGCGCATAG 1627  
Qy 554 ACCAGATCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 613  
Db 1628 ACCAGATCTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCCCA 1687  
Qy 614 CCCCCTGAAATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCAACCTTGAATTC 673  
Db 1688 CCCCCTGAAATCATCGGCCGCAACATGCTGACCCAGCTGGGCTGCAACCTTGAATTC 1747  
Qy 674 CCATCAGCCCCATCAGACCGTGGCGTGAAGCTGAAGCCCGGATGAAGCCCGCAAGG 733  
Db 1748 CCATCAGCCCCATCAGACCGTGGCGTGAAGCTGAAGCCCGGATGAAGCCCGCAAGG 1807  
Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGCAGGAGA 793  
Db 1808 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGCAGGAGA 1867  
Qy 794 TCGAGAAGAGGCGCAAGATCAACAGATCGGCCCGAGAACCCCTTACACACCCCGTGT 853  
Db 1868 TCGAGAAGAGGCGCAAGATCAACAGATCGGCCCGAGAACCCCTTACACACCCCGTGT 1927  
Qy 854 TCGCCATCAAGAAGAGNACAGACCAAGTGGCGCAAGCTGTGACCTTCGCGAGCTGA 913  
Db 1928 TCGCCATCAAGAAGAGNACAGACCAAGTGGCGCAAGCTGTGACCTTCGCGAGCTGA 1987  
Qy 914 ACAAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCATCCCGACCCCGCGCCTGA 973  
Db 1988 ACAAGCGCACCCAGACTTCTGGGAGGTGACGTGGGCATCCCGACCCCGCGCCTGA 2047  
Qy 974 AGAAGAAGAGCGGTGACCGTGTGACGTGGGCGAGCCCTTACAGCTGCGCCCTGG 1033  
Db 2048 AGAAGAAGAGCGGTGACCGTGTGACGTGGGCGAGCCCTTACAGCTGCGCCCTGG 2107  
Qy 1034 ACGAGACTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAACACGAGACCCCG 1093  
Db 2108 ACGAGACTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAACACGAGACCCCG 2167  
Qy 1094 GCATCCGCTACCAGTACACCGTCTGCCCGAGGCTGGAAGGCGAGCCCGACATCTCC 1153  
Db 2168 GCATCCGCTACCAGTACACCGTCTGCCCGAGGCTGGAAGGCGAGCCCGACATCTCC 2227  
Qy 1154 AGACAGCATGACCAAGATCTCGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 1213  
Db 2228 AGACAGCATGACCAAGATCTCGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCT 2287

Qy 1214 ACCAGGCCCTTGTATCGTGGGCGAGCACTGGAGATCGGCAGCACCGCGCAAGATCG 1273  
Db 2288 ACCAGGCCCTTGTATCGTGGGCGAGCACTGGAGATCGGCAGCACCGCGCAAGATCG 2347  
Qy 1274 AGGAGCTGCGCAAGCACTGTGCGCTGGGCTTTCACACCCCGCAAGAAGACCAAGA 1333  
Db 2348 AGGAGCTGCGCAAGCACTGTGCGCTGGGCTTTCACACCCCGCAAGAAGACCAAGA 2407  
Qy 1334 AGGAGCCCTTCTGCTGCGCATCGAGTGCACCCGCAAGTGAACCTGTGAGCCGTGAGCCCATCG 1393  
Db 2408 AGGAGCCCTTCTGCTGCGCATCGAGTGCACCCGCAAGTGAACCTGTGAGCCCATCG 2467  
Qy 1394 AGCTGCCGAGAAAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGA 1453  
Db 2468 AGCTGCCGAGAAAGAGAGCTGGACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGA 2527  
Qy 1454 ACTGGGCGAGCAGATCTTACCCCGGCATCAAGGTGCGCGCATCAAGGTGCTGCGCG 1513  
Db 2528 ACTGGGCGAGCAGATCTTACCCCGGCATCAAGGTGCGCGCATCAAGGTGCTGCGCG 2587  
Qy 1514 GCGCAAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCG 1573  
Db 2588 GCGCAAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCG 2647  
Qy 1574 AGAACCGGAGATCTGCGCGAGCCGTGACCGCGTGTATCTACGACCCCGCAAGGACC 1633  
Db 2648 AGAACCGGAGATCTGCGCGAGCCGTGACCGCGTGTATCTACGACCCCGCAAGGACC 2707  
Qy 1634 TGGTGGCGAGATCCAGAAGCAGGCGCCACGACCAAGTGAACCTTACAGATCTTACAGGAGC 1693  
Db 2708 TGGTGGCGAGATCCAGAAGCAGGCGCCACGACCAAGTGAACCTTACAGATCTTACAGGAGC 2767  
Qy 1694 CTTTCAAGAACCTGGAAGACCGCAAGTAGCCAGATGGCCAGATGGCCACCGCCACCAAGCAGG 1753  
Db 2768 CTTTCAAGAACCTGGAAGACCGCAAGTAGCCAGATGGCCAGATGGCCACCGCCACCAAGCAGG 2827  
Qy 1754 TGAAGCAGCTGACCGAGGCGGTGCAAGAGATGCCATGGAGAGCATCGTGATCTGGGCA 1813  
Db 2828 TGAAGCAGCTGACCGAGGCGGTGCAAGAGATGCCATGGAGAGCATCGTGATCTGGGCA 2887  
Qy 1814 AGACCCCAAGTTCGCTGCGCATCCAGAAGGAGACCTGGGAGACCTGGTGAACCGACT 1873  
Db 2888 AGACCCCAAGTTCGCTGCGCATCCAGAAGGAGACCTGGGAGACCTGGTGAACCGACT 2947  
Qy 1874 ACTGCGCGCACCTGAGTCCCGAGTGGGAGTTGCTGAACACCCCGCCCTGGTGAAGC 1933  
Db 2948 ACTGCGCGCACCTGAGTCCCGAGTGGGAGTTGCTGAACACCCCGCCCTGGTGAAGC 3007  
Qy 1934 TGTGTTACAGCTGGAGAAGGAGCCATCATCGGCGCGAGACCTTCTACGTGGACCGCG 1993  
Db 3008 TGTGTTACAGCTGGAGAAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGACCGCG 3067  
Qy 1994 CGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGCGCGGCA 2053  
Db 3068 CGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGCGCGGCA 3127  
Qy 2054 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGG 2113  
Db 3128 AGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGGCGCATCCAGCTGG 3187  
Qy 2114 CCCTGCGAGACGCGGCGAGCGAGTGAACATCGTGACCGAGCAGCTACGCCCTTGGGCA 2173  
Db 3188 CCCTGCGAGACGCGGCGAGCGAGTGAACATCGTGACCGAGCAGCTACGCCCTTGGGCA 3247  
Qy 2174 TCATCCAGGCCCGCAGCCGCAAGAGCGAGAGCTGGTGAACCAAGATCATCGAGAGC 2233  
Db 3248 TCATCCAGGCCCGCAGCCGCAAGAGCGAGAGCTGGTGAACCAAGATCATCGAGAGC 3307  
Qy 2234 TCATCCAGAGGAGAGAGTGTACTGAGCTGGGTGCGGCCCAAGGGCATCGGGGCA 2293  
Db 3308 TCATCCAGAGGAGAGAGTGTACTGAGCTGGGTGCGGCCCAAGGGCATCGGGGCA 3367

Qy 2294 ACAGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATCCCGCAAGGTGCTGTCTTGAGCGGCA 2353  
Db |||||  
3368 ACAGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATCCCGCAAGGTGCTGTCTTGAGCGGCA 3427  
Qy |||||  
2354 TCGATGGCGGCATCGTGATCTTACCAAGTACATGAGGACCTGTACGTGGGCGAGCGCGGCC 2413  
Db |||||  
3428 TCGATGGCGGCATCGTGATCTTACCAAGTACATGAGGACCTGTACGTGGGCGAGCGCGGCC 3487  
Qy |||||  
2414 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2451  
Db |||||  
3488 CTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 3525

RESULT 14  
US-10-435-45  
; Sequence 45, Application US/10190435  
; Publication No. US20030143248A1  
; GENERAL INFORMATION:  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: BARNETT, Susan W.  
; APPLICANT: LIAN, Ying  
; APPLICANT: ENGELBRECHT, Susan  
; APPLICANT: VAN RENSBERG, Esrelita J.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING ANTIGENIC HIV TYPE C  
; FILE REFERENCE: PP18133.003 / 2302-18133  
; CURRENT APPLICATION NUMBER: US/10/190,435  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 319  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 45  
; LENGTH: 2457  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: p2Polopt\_c  
US-10-435-45

Query Match 97.8%; Score 2403.4; DB 15; Length 2457;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 2439; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

Qy 7 GCCACCATGGCCGAGGCCATGAGCGAGCCACAGCGCCCAACATCTCTGATGAGCGAGC 66  
Db 1 GCCACCATGGCCGAGGCCATGAGCGAGCCACAGCGCCCAACATCTCTGATGAGCGAGC 60

Qy 67 AACTTCAAGGCCCCAGCGCATCATCAAGTGTCTTAACTCGCGCAAGAGGGCCACATC 126  
Db 61 AACTTCAAGGCCCCAGCGCATCATCAAGTGTCTTAACTCGCGCAAGAGGGCCACATC 120

Qy 127 GCCGCAACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGCCAC 186  
Db 121 GCCGCAACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGCCAC 180

Qy 187 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCC 246  
Db 181 CAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCC 240

Qy 247 CAGGCGAAGGCCCGCGAGTTCCCCCAGGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 306  
Db 241 CAGGCGAAGGCCCGCGAGTTCCCCCAGGAGCAGAAACCGCGCAACAGCCCCACAGCGCG 300

Qy 307 GAGCTGAGGTGGCGGCGCAACCCCGAGCGAGCGCGCGCGCGCGAGCGCGCGAGCGCGCACC 366  
Db 301 GAGCTGAGGTGGCGGCGCAACCCCGAGCGAGCGCGCGCGCGCGCGAGCGCGCGAGCGCGCACC 360

Qy 367 CTGAACCTTCCCCCAGATCACCTGTGTGAGCGCGCCCTCTGTGTGAGCATCAAGGTGGCGGC 426  
Db 361 CTGAACCTTCCCCCAGATCACCTGTGTGAGCGCGCCCTCTGTGTGAGCATCAAGGTGGCGGC 420

Qy 427 CAGATCAAGGAGGCCCTGTGTGAACCCGGGCGCGAGCACACCGTGTGTGAGGAGATGAGC 486  
Db 421 CAGATCAAGGAGGCCCTGTGTGAACCCGGGCGCGAGCACACCGTGTGTGAGGAGATGAGC 480

Qy 487 CTGCCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGGTGGCG 546  
Db |||||  
481 CTGCCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGGCTTTCATCAAGGTGGCG 540  
Qy |||||  
547 CAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGACCGTGTGATC 606  
Db |||||  
541 CAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGACCGTGTGATC 600  
Qy |||||  
607 GGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGTG 666  
Db |||||  
601 GGCCCCACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCGTG 660  
Qy |||||  
667 AACTTCCCCCATCGCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCGCATGGACGGC 726  
Db |||||  
661 AACTTCCCCCATCGCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCGCATGGACGGC 720  
Qy |||||  
727 CCAAGGTGAAGCAGTGGCCCCCTGACCGAGAGAGATCAAGGCCCTTGACCGCCATCTGC 786  
Db |||||  
721 CCAAGGTGAAGCAGTGGCCCCCTGACCGAGAGAGATCAAGGCCCTTGACCGCCATCTGC 780  
Qy |||||  
787 GAGGAGATGAGAGAGGAGGCGCAGATCACCAAGATCGGCCCGCGAGAACCCCTCAACACC 846  
Db |||||  
781 GAGGAGATGAGAGAGGAGGCGCAGATCACCAAGATCGGCCCGCGAGAACCCCTCAACACC 840  
Qy |||||  
847 CCGCTGTTCCGCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 906  
Db |||||  
841 CCGCTGTTCCGCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGC 900  
Qy |||||  
907 GAGCTGAACCAAGCGACCCAGGACTTCTGGAGAGTGCAGTGGGCATCCCCCACCCTCCGCC 966  
Db |||||  
901 GAGCTGAACCAAGCGACCCAGGACTTCTGGAGAGTGCAGTGGGCATCCCCCACCCTCCGCC 960  
Qy |||||  
967 GGCTGAAGAGAGAGAGGCTGACCGTGTGACGCTGGCGAGCGCTACTTTCAGCGTG 1026  
Db |||||  
961 GGCTGAAGAGAGAGAGGCTGACCGTGTGACGCTGGCGAGCGCTACTTTCAGCGTG 1020  
Qy |||||  
1027 CCGCTGAAGAGAGAGGACTTCCGCATGACACCGCTTACACCTCCCGAGCATCAACAAACGAG 1086  
Db |||||  
1021 CCGCTGAAGAGAGGACTTCCGCATGACACCGCTTACACCTCCCGAGCATCAACAAACGAG 1080  
Qy |||||  
1087 ACCCCCGGCGATCCGCTACCAAGTACAACTGTGCCCCAGGGCTGGAAAGGCGAGCCCCCAGC 1146  
Db |||||  
1081 ACCCCCGGCGATCCGCTACCAAGTACAACTGTGCCCCAGGGCTGGAAAGGCGAGCCCCCAGC 1140  
Qy |||||  
1147 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCGCAACCCCGAGATC 1206  
Db |||||  
1141 ATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCGCAACCCCGAGATC 1200  
Qy |||||  
1207 GTGATCTACCA-----GGCCCCCTCTGTAGTGGGCGAGCGACCTGGAGATCGGCGCAGCAC 1260  
Db |||||  
1201 GTGATCTACCAAGTACATGAGAGCCTGTAGTGGGCGAGCGACCTGGAGATCGGCGCAGCAC 1260  
Qy |||||  
1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCATCTGTGCGCTGGGGCTTTCACCAACCCCGGAC 1320  
Db |||||  
1261 CGCGCCAAGATCGAGGAGCTGCGCAAGCATCTGTGCGCTGGGGCTTTCACCAACCCCGGAC 1320  
Qy |||||  
1321 AAGAGCACAGAAAGAGGCCCTTCTCTGCCCT-----CGAGCTGCACCCCGAGCAG 1374  
Db |||||  
1321 AAGAGCACAGAAAGAGGCCCTTCTCTGCCCTGGGGCTACGAGCTGCACCCCGAGCAG 1380  
Qy |||||  
1375 TGGACCGTGCAGCCATCGAGCTGCGCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1434  
Db |||||  
1381 TGGACCGTGCAGCCATCGAGCTGCGCGAGAGAGAGCTGGACCGGTGAACGACATCCAG 1440  
Qy |||||  
1435 AAGTGTGTGGCGCAAGCTGAATCTGGCGCAGCAGATCTACCCCGGCGCATCAAGGTGCGCCAG 1494  
Db |||||  
1441 AAGTGTGTGGCGCAAGCTGAATCTGGCGCAGCAGATCTACCCCGGCGCATCAAGGTGCGCCAG 1500  
Qy |||||  
1495 CTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTGCCCTTGAACCGAGGAG 1554  
Db |||||  
1501 CTGTGCAAGCTGTGCGCGGCGCAAGGCCCTGACCGACATCTGTGCCCTTGAACCGAGGAG 1560



Db 841 CCCGTGTTCCCATCAAGAGAGGACAGCACCAGTGGCGCAAGCTGGTGGACTTCCGC 900  
Qy 907 GAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGATCCCCACCCCGCC 966  
Db 901 GAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGATCCCCACCCCGCC 960  
Qy 967 GGCCTGAAGAGAGAGAGGCTGACCGTGTGAGAGTGGGCGAGCGCTACTTCAAGCGTG 1026  
Db 961 GGCCTGAAGAGAGAGAGGCTGACCGTGTGAGAGTGGGCGAGCGCTACTTCAAGCGTG 1020  
Qy 1027 CCCCTGGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAGAG 1086  
Db 1021 CCCCTGGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAGAG 1080  
Qy 1087 ACCCCCGGACCTCCGCTACCACTACAAAGTGTGCTGCCCAAGGGCTGGAAGGGCAGCCCGAGC 1146  
Db 1081 ACCCCCGGACCTCCGCTACCACTACAAAGTGTGCTGCCCAAGGGCTGGAAGGGCAGCCCGAGC 1140  
Qy 1147 ATCTTCCAGAGAGATGACCAAGATCTGTGAGCGCTTCCGGCGCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCCAGAGAGATGACCAAGATCTGTGAGCGCTTCCGGCGCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACCA- ----GCCCCCTGTGACCTGGGCGAGCGACTGGAGATCGGCCAGC 1260  
Db 1201 GTGATCTACCACTACCACTGACCTGTGAGCGAGCGACTGGAGATCGGCCAGC 1260  
Qy 1261 CGCGCCAAGATCGAGGAGCTCGCAAGCACTGTGCTGGGCTTCAACCAACCCCGAGC 1320  
Db 1261 CGCGCCAAGATCGAGGAGCTCGCAAGCACTGTGCTGGGCTTCAACCAACCCCGAGC 1320  
Qy 1321 AAGAAGCAACAGAGAGCGCCCTTCTGCTGCCAT- ----CGAGCTGCACCCCGCAAG 1374  
Db 1321 AAGAAGCAACAGAGAGCGCCCTTCTGCTGGATGGGCTACGAGCTGCACCCCGCAAG 1380  
Qy 1375 TGGACCTGTGAGCCATCGAGCTGCCGAGAGGAGCTGGAAGCTGGAAGCATCCAG 1434  
Db 1381 TGGACCTGTGAGCCATCGAGCTGCCGAGAGGAGCTGGAAGCTGGAAGCATCCAG 1440  
Qy 1435 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCGAGATCTACCCCGGCTCAAGGTGGCGAG 1494  
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCGCAGCGAGATCTACCCCGGCTCAAGGTGGCGAG 1500  
Qy 1495 CTGTGCAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCTGTGCCCTGACCGAGGAG 1554  
Db 1501 CTGTGCAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCTGTGCCCTGACCGAGGAG 1560  
Qy 1555 GCGAGCTGGAGCTGGCGCGAAGCGGAGATCTGCGGAGCGCGTGCAGCGGCTGTAC 1614  
Db 1561 GCGAGCTGGAGCTGGCGCGAAGCGGAGATCTGCGGAGCGCGTGCAGCGGCTGTAC 1620  
Qy 1615 TAGACCCCGAGCAAGGAGCTGGTGGCGGAGATCCAGAGAGGCGCCAGCAGTGGACC 1674  
Db 1621 TAGACCCCGAGCAAGGAGCTGGTGGCGGAGATCCAGAGAGGCGCCAGCAGTGGACC 1680  
Qy 1675 TACCAGATCTACAGGAGCGCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGATGCGC 1734  
Db 1681 TACCAGATCTACAGGAGCGCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGATGCGC 1740  
Qy 1735 ACCGCCACACCAAGAGTGAAGAGCTGACCGAGCGCGCTGCAAGAGATCGCCATGGAG 1794  
Db 1741 ACCGCCACACCAAGAGTGAAGAGCTGACCGAGCGCGCTGCAAGAGATCGCCATGGAG 1800  
Qy 1795 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTTCCAGAGGAGAGCTGG 1854  
Db 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTTCCAGAGGAGAGCTGG 1860  
Qy 1855 GAGACCTGGTGGACCGACTACTTGGCAGCGCACCTGGATCCCGGAGTGGGATTCGTGAAC 1914  
Db 1861 GAGACCTGGTGGACCGACTACTTGGCAGCGCACCTGGATCCCGGAGTGGGATTCGTGAAC 1920  
Qy 1915 ACCCCCCCTGGTGAAGCTGTGTATCCAGCTGAGAGAGAGCCCATCATCGGCGCGAG 1974  
Db 1921 ACCCCCCCTGGTGAAGCTGTGTATCCAGCTGAGAGAGAGCCCATCATCGGCGCGAG 1980

Search completed: June 2, 2005, 04:23:48  
Job time : 1415.9 secs

Qy 1975 ACCTTTCTACGTGGACGGCGCCCAACCGCGAGACCAGATCGGCAAGCGCGGTACGTG 2034  
Db 1981 ACCTTTCTACGTGGACGGCGCCCAACCGCGAGACCAGATCGGCAAGCGCGGTACGTG 2040  
Qy 2035 ACCGACCGGGCGCGGAGAGATCGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2094  
Db 2041 ACCGACCGGGCGCGGAGAGATCGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2100  
Qy 2095 CTGCAGGCCATCCAGCTGGCCCTCAGGACAGCGGCGAGCGAGGTGAACATCTGTACCGAC 2154  
Db 2101 CTGCAGGCCATCCAGCTGGCCCTCAGGACAGCGGCGAGCGAGGTGAACATCTGTACCGAC 2160  
Qy 2155 AGCCAGTACGCCCTTGGGCATCATCCAGGCCAGGCCCGACAAGAGCGAGCGAGTGGTG 2214  
Db 2161 AGCCAGTACGCCCTTGGGCATCATCCAGGCCAGGCCCGACAAGAGCGAGCGAGTGGTG 2220  
Qy 2215 AACAGATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2274  
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2280  
Qy 2275 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2334  
Db 2281 CACAAGGGCATCGGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG 2340  
Qy 2335 GTGCTGTTCTTGGACGGCATCGATGGCGGCATCTGATCTACCAAGTACATGAGACGCTG 2394  
Db 2341 GTGCTGTTCTTGGACGGCATCGATGGCGGCATCTGATCTACCAAGTACATGAGACGCTG 2400  
Qy 2395 TACGTGGGCGAGCGCGCGCTTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2451  
Db 2401 TACGTGGGCGAGCGCGCGCTTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACCGGT 2457

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2005, 23:20:00 ; Search time 1267.33 Seconds  
(without alignments)  
11504.726 Million cell updates/sec

Title: US-09-610-313b-31  
Perfect score: 2463  
Sequence: 1 gtcagccaccatgcccga.....gggctagaccggtgaattc 2463

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2463	100.0	2463	6 ABL39960	ABL39960 Synthetic
2	2463	100.0	2463	12 ADM73765	Adm73765 HIV-1 pol
3	2455.4	99.7	2457	8 ACA03547	ACA03547 Synthetic
4	2455.4	99.7	2457	10 ADC13265	Adc13265 DNA of HI
5	2442.2	99.2	2469	6 ABL39959	ABL39959 Synthetic
6	2442.2	99.2	2469	12 ADM73764	Adm73764 HIV-1 pol
7	2436.2	98.9	2457	6 ABL39961	ABL39961 Synthetic
8	2436.2	98.9	2457	12 ADM73766	Adm73766 HIV-1 pol
9	2430.2	98.7	2457	8 ACA03548	ACA03548 Synthetic
10	2430.2	98.7	2457	10 ADC13266	Adc13266 DNA of HI
11	2422.6	98.4	2445	8 ACA03546	ACA03546 Synthetic
12	2422.6	98.4	2445	10 ADC13264	Adc13264 DNA of HI
13	2415.6	98.1	3930	10 ADC13230	Adc13230 DNA of HI
14	2414	98.0	3930	10 ADC13231	Adc13231 DNA of HI
15	2414	98.0	3930	10 ADC13232	Adc13232 DNA of HI
16	2414	98.0	5184	8 ACA03591	ACA03591 Synthetic
17	2414	98.0	5184	10 ADC13279	Adc13279 DNA of HI
18	2383.6	96.8	3531	10 ADC13234	Adc13234 DNA of HI
19	2382	96.7	3537	10 ADC13236	Adc13236 DNA of HI
20	2381	96.7	5145	8 ACA03521	ACA03521 Synthetic

## ALIGNMENTS

### RESULT 1

ABL39960 ID ABL39960 standard; DNA; 2463 BP.

XX ABL39960;

XX 15-MAY-2002 (first entry)

DE Synthetic construct PR975YM SEQ ID NO:31.

XX Human immunodeficiency virus type C; antigenic HIV type C protein;  
immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; env; nef;  
immunostimulant; gene therapy; gene; ds.

OS Human immunodeficiency virus; type C.  
OS Synthetic.

PN WO200204493-A2.

XX 17-JAN-2002.

XX 05-JUL-2001; 2001WO-US021241.

XX 05-JUL-2000; 2000US-00610313.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2002-154920/20.

XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in applications including DNA immunization or generation of packaging cell lines, particularly in gene therapy.

XX Claim 1; Fig 9; 233pp; English.

CC The present invention describes expression cassettes comprising a polynucleotide sequence encoding a polypeptide comprising immunogenic HIV type C polypeptides. The expression cassettes comprise any of the HIV type C sequences encoding gag, pol, vif, vpr, tat, rev, env or nef (i). (i) have immunostimulant activity and can be used in gene therapy.

21 2381 96.7 5145 10 ADC13233 Adc13233 DNA of HI  
22 2371 96.3 3538 10 ADC13235 Adc13235 DNA of HI  
23 2322.6 94.3 3607 8 ACA03551 Aca03551 Synthetic  
24 2322.6 94.3 3607 10 ADC13269 Adc13269 DNA of HI  
25 2322.6 94.3 3624 8 ACA03550 Aca03550 Synthetic  
26 2322.6 94.3 3624 10 ADC13268 Adc13268 DNA of HI  
27 2304.4 93.6 3597 8 ACA03549 Aca03549 Synthetic  
28 2304.4 93.6 3597 10 ADC13267 Adc13267 DNA of HI  
29 2142 87.0 2466 8 ACA03542 Aca03542 Synthetic  
30 2142 87.0 2466 8 ACC78506 Acc78506 HIV p2Pol  
31 2122.8 86.2 2472 8 ACA03543 Aca03543 Synthetic  
32 2122.8 86.2 2472 8 ACC78507 Acc78507 HIV p2Pol  
33 2115.2 85.9 2460 8 ACA03541 Aca03541 Synthetic  
34 2115.2 85.9 2460 8 ACC78505 Acc78505 HIV p2Pol  
35 2114.2 85.8 3564 8 ACC78488 Acc78488 HIV GagPo  
36 2114.2 85.8 3564 8 ACC78489 Acc78489 HIV GagPo  
37 2113.6 85.8 4716 8 ACA03522 Aca03522 Synthetic  
38 2113.6 85.8 4716 10 ADC13238 Adc13238 DNA of HI  
39 2109.8 85.7 3999 8 ACC78484 Acc78484 HIV GagCo  
40 2108.2 85.6 3999 8 ACC78485 Acc78485 HIV GagCo  
41 2108.2 85.6 3999 8 ACC78486 Acc78486 HIV GagCo  
42 2108.2 85.6 5283 8 ACA03584 Aca03584 Synthetic  
43 2108.2 85.6 5283 8 ACC78529 Acc78529 HIV TatRe  
44 2108 85.6 4713 8 ACA03592 Aca03592 Synthetic  
45 2108 85.6 4713 10 ADC13280 Adc13280 DNA of HI

CC	The HIV type C polynucleotides are useful in applications including DNA									
CC	immunisation, generation of packaging cell lines, and production of HIV									
CC	type C proteins. The polynucleotides are particularly useful in gene									
CC	therapy and DNA immunisation applications. ABL39942 to ABL40054 and									
CC	ABB06204 to ABB06215 represent sequences used in the exemplification of									
CC	the present invention									
XX										
SQ	Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;									
	Query Match 100.0%; Score 2463; DB 6; Length 2463;									
	Best Local Similarity 100.0%; Pred. No. 3.9e-297;									
	Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	GTGAGCGCCACCATGCGCGGCGCATGAGCGAGCCACGAGCGCCACATCTCTGATGCAG	60							
Db	1	GTGAGCGCCACCATGCGCGGCGCATGAGCGAGCCACGAGCGCCACATCTCTGATGCAG	60							
Qy	61	CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTTCAACTGCGGCAAGGAGGGC	120							
Db	61	CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTTCAACTGCGGCAAGGAGGGC	120							
Qy	121	CACATCGCCGCAACTCGCCGCGCCCGCCGCAAGAAGGGCTGTGGAAGTGCGCCAAGGAG	180							
Db	121	CACATCGCCGCAACTCGCCGCGCCCGCCGCAAGAAGGGCTGTGGAAGTGCGCCAAGGAG	180							
Qy	181	GGCACCAGATGAAGACTATGACCGAGCGCGCAGGCCAACTTCTTCGCGAGGACTGGCC	240							
Db	181	GGCACCAGATGAAGACTATGACCGAGCGCGCAGGCCAACTTCTTCGCGAGGACTGGCC	240							
Qy	241	TTCCCCAGGGCAAGCCCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACACGCCCAACC	300							
Db	241	TTCCCCAGGGCAAGCCCGCGAGTTCCTCCAGCGAGCAGAACCGCGCAACACGCCCAACC	300							
Qy	301	AGCGCGAGTGCAGGTGCGCGCGCAACAACCCCGCAGCGAGGCCCGCGCGCGAGCCAG	360							
Db	301	AGCGCGAGTGCAGGTGCGCGCGCAACAACCCCGCAGCGAGGCCCGCGCGCGAGCCAG	360							
Qy	361	GGCACCCTGAACTTCCCGCCAGATCACCTGTGTCAGGCGCCCTCTGTGAGCATCAAGGTG	420							
Db	361	GGCACCCTGAACTTCCCGCCAGATCACCTGTGTCAGGCGCCCTCTGTGAGCATCAAGGTG	420							
Qy	421	GGCGGCGAGATCAAGAGGCGCTGTGGACACCGGCGCGCAGACACCGTGTGGAGGAG	480							
Db	421	GGCGGCGAGATCAAGAGGCGCTGTGGACACCGGCGCGCAGACACCGTGTGGAGGAG	480							
Qy	481	ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGGCGGCATCGCGCGCTTTCATCAAG	540							
Db	481	ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGGCGGCATCGCGCGCTTTCATCAAG	540							
Qy	541	GTGCGCGAGTACGACCAAGATCTTGATCGAGATCTGCGGCAGAGAGGCCATCGGACCGTG	600							
Db	541	GTGCGCGAGTACGACCAAGATCTTGATCGAGATCTGCGGCAGAGAGGCCATCGGACCGTG	600							
Qy	601	CTGATCGGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660							
Db	601	CTGATCGGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660							
Qy	661	ACCCTGAACTTCCCTCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATG	720							
Db	661	ACCCTGAACTTCCCTCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATG	720							
Qy	721	GACGGCCCAAGTGAAGCGTGGCCCTGTACCGAGAGAGATCAAGGCCCTGACCGCC	780							
Db	721	GACGGCCCAAGTGAAGCGTGGCCCTGTACCGAGAGAGATCAAGGCCCTGACCGCC	780							
Qy	781	ATCTGCGAGGAGTGAAGAGGAGGCAAGATCAACAAGATCGGCCCGCGAGAACCCCTAC	840							
Db	781	ATCTGCGAGGAGTGAAGAGGAGGCAAGATCAACAAGATCGGCCCGCGAGAACCCCTAC	840							
Qy	841	AACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCGCAAGCTGGTGAC	900							
Db	841	AACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCGCAAGCTGGTGAC	900							

Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCCAC	960							
Db	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCCAC	960							
Qy	961	CCCGCGGCGCTGAAGAAGAAGAGAGCGTGCACCGTGTGACGCTGGCGCGACCGCTACTTC	1020							
Db	961	CCCGCGGCGCTGAAGAAGAAGAGAGCGTGCACCGTGTGACGCTGGCGCGACCGCTACTTC	1020							
Qy	1021	AGCGTGCCTTGAACGAGGACTTCCGCAAGTACACCGCTTACCATCTCCGAGATCAAC	1080							
Db	1021	AGCGTGCCTTGAACGAGGACTTCCGCAAGTACACCGCTTACCATCTCCGAGATCAAC	1080							
Qy	1081	AACGAGACCCCGGCGATCCGCTACCAAGTACAAAGTGTGCTGCCAGGCTGGAAGGGCAGC	1140							
Db	1081	AACGAGACCCCGGCGATCCGCTACCAAGTACAAAGTGTGCTGCCAGGCTGGAAGGGCAGC	1140							
Qy	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGGCCCTTCCGCGCGCCCAACCCC	1200							
Db	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGAGGCCCTTCCGCGCGCCCAACCCC	1200							
Qy	1201	GAGATCGTATCTACAGGCGCCCTGTAGCTGGGAGAGACCTGAGATCGGCGAGCAC	1260							
Db	1201	GAGATCGTATCTACAGGCGCCCTGTAGCTGGGAGAGACCTGAGATCGGCGAGCAC	1260							
Qy	1261	CGCGCAAGATCGAGGAGCTGCGAAGCACCTGCTGCGTGGGGCTTACCAACCCCGAC	1320							
Db	1261	CGCGCAAGATCGAGGAGCTGCGAAGCACCTGCTGCGTGGGGCTTACCAACCCCGAC	1320							
Qy	1321	AAGAAGCACAGAAAGAGCGCCCTTCTGTGTGATGGGCTTACGAGGTGCACCCCGCAAG	1380							
Db	1321	AAGAAGCACAGAAAGAGCGCCCTTCTGTGTGATGGGCTTACGAGGTGCACCCCGCAAG	1380							
Qy	1381	TGACCGTGCAGCCATCGAGCTGCCGAGAAAGAGAGTGGACCGTGAACGATCCAG	1440							
Db	1381	TGACCGTGCAGCCATCGAGCTGCCGAGAAAGAGAGTGGACCGTGAACGATCCAG	1440							
Qy	1441	AAGCTGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGCATCAAGGTGCGCAG	1500							
Db	1441	AAGCTGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCCGCATCAAGGTGCGCAG	1500							
Qy	1501	CTGTGAAGCTGTGCGCGCGCCCAAGGGCTTACCGCATCTGTCCTTACCGAGGAG	1560							
Db	1501	CTGTGAAGCTGTGCGCGCGCCCAAGGGCTTACCGCATCTGTCCTTACCGAGGAG	1560							
Qy	1561	GCGAGCTGAGCTGCGCGAGAACCGCGAGATCTCTGCGGAGCCCTGACCGCGGTGAC	1620							
Db	1561	GCGAGCTGAGCTGCGCGAGAACCGCGAGATCTCTGCGGAGCCCTGACCGCGGTGAC	1620							
Qy	1621	TACGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGGCCACGACAGTGGACC	1680							
Db	1621	TACGACCCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGGCCACGACAGTGGACC	1680							
Qy	1681	TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGC	1740							
Db	1681	TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGC	1740							
Qy	1741	ACCGCCCAACCAACGAGTGAAGAGTGAACCGAGGCGGTGAGAGATGCGCATGGAG	1800							
Db	1741	ACCGCCCAACCAACGAGTGAAGAGTGAACCGAGGCGGTGAGAGATGCGCATGGAG	1800							
Qy	1801	AGCATCTGTGGGCAAGACCCCAAGTTCGGCTGCGCCATCAGAGAGGAGACTGG	1860							
Db	1801	AGCATCTGTGTGGGCAAGACCCCAAGTTCGGCTGCGCCATCAGAGAGGAGACTGG	1860							
Qy	1861	GAGACCTGTGGAGCCACTACTTGGCAGGCCACCTGGATCCCCGAGTGGAGTTCGTGAAC	1920							
Db	1861	GAGACCTGTGTGGAGCCACTACTTGGCAGGCCACCTGGATCCCCGAGTGGAGTTCGTGAAC	1920							
Qy	1921	ACCCCCCTGTGTGAAGCTGTGTTACGAGTGGAGAGGCCCATCATCGCGCGCGAG	1980							
Db	1921	ACCCCCCTGTGTGAAGCTGTGTTACGAGTGGAGAGGCCCATCATCATCGCGCGCGAG	1980							
Qy	1981	ACCTTCTACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACGTG	2040							

Db 1981 ACCCTTCTACGTGGACGGCGCCGCCAACCCGCGAGACAAGATCGGCAAGCGCGGTACGTG 2040  
 Qy 2041 ACCGACCGGGCGCGCAGAGATCGTGAAGCTGAGCCGACACCAACAGAGACCGAG 2100  
 Db 2041 ACCGACCGGGCGCGCAGAGATCGTGAAGCTGAGCCGACACCAACAGAGACCGAG 2100  
 Qy 2101 CTGCAAGGCATCCAGCTGGCCCTGAGACAGCGGCGAGGTTGAACATCGTGACCGAC 2160  
 Db 2101 CTGCAAGGCATCCAGCTGGCCCTGAGACAGCGGCGAGGTTGAACATCGTGACCGAC 2160  
 Qy 2161 AGCCAGTACGCTGGCGATCATCCAGGCCAGCCGACAGAGCGAGCGAGCTGTG 2220  
 Db 2161 AGCCAGTACGCTGGCGATCATCCAGGCCAGCCGACAGAGCGAGCGAGCTGTG 2220  
 Qy 2221 AACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGAGGTGATCTGAGCTGGGTGCCGCC 2280  
 Db 2221 AACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGAGGTGATCTGAGCTGGGTGCCGCC 2280  
 Qy 2281 CACAAAGGCATCGCGCGCAACAGAGCAGATCGAAGCTGTGAGCAAGGGGATCCGCAAG 2340  
 Db 2281 CACAAAGGCATCGCGCGCAACAGAGCAGATCGAAGCTGTGAGCAAGGGGATCCGCAAG 2340  
 Qy 2341 GTGCTGTTCCTGGACGGCATCGATCGCGGCATCGTGATCTACCAATGATGACGACCTG 2400  
 Db 2341 GTGCTGTTCCTGGACGGCATCGATCGCGGCATCGTGATCTACCAATGATGACGACCTG 2400  
 Qy 2401 TACGTGGCGAGCGCGCGCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
 Db 2401 TACGTGGCGAGCGCGCGCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
 Qy 2461 TTC 2463  
 Db 2461 TTC 2463

RESULT 2  
 ID ADM73765 standard; DNA; 2463 BP.  
 AC ADM73765;  
 XX  
 DT 03-JUN-2004 (first entry)  
 XX  
 DE HIV-1 polynucleotide #8.  
 XX  
 KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;  
 XX  
 KW HIV type C protein; immunostimulant.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 OS US2003223961-A1.  
 XX  
 PD 04-DEC-2003.  
 XX  
 PF 05-JUL-2001; 2001US-00899575.  
 XX  
 PR 05-JUL-2000; 2000US-00610313.  
 XX  
 PA (MEGE/) MEGEDE J Z.  
 PA (BARN/) BARNETT S W.  
 PA (ENGE/) ENGELBRECHT S.  
 XX (RENS/) RENSBURG E J V.  
 PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;  
 XX WPI; 2004-060515/06.  
 DR  
 XX  
 XX  
 PT New expression cassette comprising a polynucleotide sequence encoding an  
 PT HIV pol polypeptide, useful in eliciting an immune response, in DNA  
 PT immunization, generating of packaging cell lines or in producing HIV Type  
 PT C proteins.  
 XX

PS Claim 1; SEQ ID NO 31; 160pp; English.  
 XX  
 CC The invention relates to an expression cassette comprising a  
 CC polynucleotide sequence encoding an HIV pol polypeptide. The invention  
 CC also relates to a recombinant expression system for use in a host cell  
 CC comprising an expression cassette, where the polynucleotide sequence  
 CC further comprises control elements capable of driving expression in the  
 CC selected host cell, a cell comprising an expression cassette where the  
 CC polynucleotide sequence further comprises control elements compatible  
 CC with the expression in the cell and a composition for generating an  
 CC immunological response, comprising an expression cassette. The expression  
 CC cassette and the methods of the invention are useful in eliciting an  
 CC immune response, in DNA immunisation, in generation of packaging cell  
 CC lines and in producing HIV Type C proteins. This sequence represents an  
 CC HIV-1 polynucleotide of the invention.  
 XX  
 SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 2463; DB 12; Length 2463;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-297;  
 Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GTGAGCGCACCATCGGCGAGGCCATGAGCCAGGCCACAGCGCCAAACATCTGTATGCAAG 60  
 Db 1 GTGAGCGCACCATCGGCGAGGCCATGAGCCAGGCCACAGCGCCAAACATCTGTATGCAAG 60  
 Qy 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGCG 120  
 Db 61 CGCAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGGAGGCG 120  
 Qy 121 CACATCGCCCGCAACTGCGCGCCCGCCCGAAGAGGGCTGCTGAAAGTGGCGCAAGGAG 180  
 Db 121 CACATCGCCCGCAACTGCGCGCCCGCCCGAAGAGGGCTGCTGAAAGTGGCGCAAGGAG 180  
 Qy 181 GCGCACCATGTAAGGACTGCAACGAGCGCCAGGGCCAACTTCTCCGCGAGGACCTGGCC 240  
 Db 181 GCGCACCATGTAAGGACTGCAACGAGCGCCAGGGCCAACTTCTCCGCGAGGACCTGGCC 240  
 Qy 241 TTCCCGCAGGCAAGGGCCCGGAGTTCCCGCAGCAGAGAGACCGCGCCCAACAGCCCCACC 300  
 Db 241 TTCCCGCAGGCAAGGGCCCGGAGTTCCCGCAGCAGAGAGACCGCGCCCAACAGCCCCACC 300  
 Qy 301 AGCGCGAGCTGCGAGGTGCGCGGCGACAACCCCGCAGCGAGGCGCGGCGCGAGCGCCAG 360  
 Db 301 AGCGCGAGCTGCGAGGTGCGCGGCGACAACCCCGCAGCGAGGCGCGGCGCGAGCGCCAG 360  
 Qy 361 GGCACCTTGAACCTTCCCGCAGATCACCCTGTGGGAGCGCCCTCTGTGAGCATCAAGGTG 420  
 Db 361 GGCACCTTGAACCTTCCCGCAGATCACCCTGTGGGAGCGCCCTCTGTGAGCATCAAGGTG 420  
 Qy 421 GCGCGCCAGATCAAGGAGGCGCTGTGGACACCGCGCGCGACGACACCGTGTGAGGAG 480  
 Db 421 GCGCGCCAGATCAAGGAGGCGCTGTGGACACCGCGCGCGACGACACCGTGTGAGGAG 480  
 Qy 481 ATGAGCTTGGCGGCAAGTGAAGCCAAAGATGATCGCGCGGATCGGCGGCTTCAATCAAG 540  
 Db 481 ATGAGCTTGGCGGCAAGTGAAGCCAAAGATGATCGCGCGGATCGGCGGCTTCAATCAAG 540  
 Qy 541 GTGCGCCAGTACGACCGACATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCCGTG 600  
 Db 541 GTGCGCCAGTACGACCGACATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGCCGTG 600  
 Qy 601 CTGATCGGCGCCCGCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 Db 601 CTGATCGGCGCCCGCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
 Qy 661 ACCCTGAACCTTCCCGCATCAGCCCGCATCGAGACCGGTGCCCGTGAAGCTGAAGCCCGCATG 720  
 Db 661 ACCCTGAACCTTCCCGCATCAGCCCGCATCGAGACCGGTGCCCGTGAAGCTGAAGCCCGCATG 720  
 Qy 721 GACGCGCCCGAGGTGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCGCTTGAACCGCC 780  
 Db 721 GACGCGCCCGAGGTGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCGCTTGAACCGCC 780







```
Db 541 GTGGCCAGTAGCAGACGATCTGATCGAGATCTCGGCAAGAGGCGCATCGGCACCGTG 600
Qy 601 CTGATCGGCCACACCCCGTGAACATCATTCGCGCCGCAACATGCTGAACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCACACCCCGTGAACATCATTCGCGCCGCAACATGCTGAACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAATCTCCCATCAGCCCATCAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720
Db 661 ACCCTGAATCTCCCATCAGCCCATCAGACCGTGCCTGGAAGCTGAAGCCCGGCATG 720
Qy 721 GACGCCCCAAAGGTGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCC 780
Db 721 GACGCCCCAAAGGTGAAGCAGTGGCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCC 780
Qy 781 ATCTCGAGGAGATGAGAAAGAGGCGCAAGATACCAAGATCGGCCCGCGAGAACCCCTAC 840
Db 781 ATCTCGAGGAGATGAGAAAGAGGCGCAAGATACCAAGATCGGCCCGCGAGAACCCCTAC 840
Qy 841 AACACCCCGGTTCGCCCATCAAGAAAGAGACACCAAGTGGCGCAAGCTGGTGGAC 900
Db 841 AACACCCCGGTTCGCCCATCAAGAAAGAGACACCAAGTGGCGCAAGCTGGTGGAC 900
Qy 901 TTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCAC 960
Db 901 TTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCAC 960
Qy 961 CCCGCGGCTGAAGAAAGAGAGCGTGACCGTGTGGAAGCGCGCTGGGCGACGCTACTTC 1020
Db 961 CCCGCGGCTGAAGAAAGAGAGCGTGACCGTGTGGAAGCGCGCTGGGCGACGCTACTTC 1020
Qy 1021 AGCGTCCCTTGGAGGAGACTTCGGGAAGTACACCGGCTTACCATCCCGAGCATCAAC 1080
Db 1021 AGCGTCCCTTGGAGGAGACTTCGGGAAGTACACCGGCTTACCATCCCGAGCATCAAC 1080
Qy 1081 AACGAGACCCCGGCTACCGTACCAAGTACCAAGTGTGCGCCGAGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCTACCGTACCAAGTGTGCGCCGAGGCTGGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGCAGATCAACAGATCTCTGGAGCCCTTCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGATCAACAGATCTCTGGAGCCCTTCGCGCCCGCAACCCC 1200
Qy 1201 GAGATCGTGATCTACAGGCCCTCTGTGAGGAGCGACCTGTGAGATCGGCGAGCAC 1260
Db 1201 GAGATCGTGATCTACAGGCCCTCTGTGAGGAGCGACCTGTGAGATCGGCGAGCAC 1260
Qy 1261 GCGCCAAAGATCGAGGAGCTGCGAAGCACCTGTGCGCTGGGCTTACCAACCCCGAC 1320
Db 1261 GCGCCAAAGATCGAGGAGCTGCGAAGCACCTGTGCGCTGGGCTTACCAACCCCGAC 1320
Qy 1321 AAGAAGCACCAAGAGGAGCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Db 1321 AAGAAGCACCAAGAGGAGCCCTTCTGTGGATGGGCTACGAGCTGCACCCCGACAAG 1380
Qy 1381 TGGACGTGACAGCCATGACGTGCGGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
Db 1381 TGGACGTGACAGCCATGACGTGCGGAGAGAGAGCTGGACCGTGAACGACATCCAG 1440
Qy 1441 AAGCTGTGGCAAGCTGAATCGGCCAGCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Db 1441 AAGCTGTGGCAAGCTGAATCGGCCAGCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500
Qy 1501 CTGTGCAAGCTGTGCGCGGCCAAGGCCCTGACGACATCTGTGCCCTGACCGAGGAG 1560
Db 1501 CTGTGCAAGCTGTGCGCGGCCAAGGCCCTGACGACATCTGTGCCCTGACCGAGGAG 1560
Qy 1561 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTGTGCGCGAGCCCGTGCAGCGGCTGTAC 1620
Db 1561 GCCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTGTGCGCGAGCCCGTGCAGCGGCTGTAC 1620
Qy 1621 TACGACCCCGAGAGGACCTGGTGGCCGAGATCAGAGCAGGCGCAGCAGCTGACGACC 1680
Db 1621 TACGACCCCGAGAGGACCTGGTGGCCGAGATCAGAGCAGGCGCAGCAGCTGACGACC 1680
```

```
Qy 1681 TACCAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGCAAGTACGCCAAGATGGC 1740
Db 1681 TACCAGATCTACAGGAGCCCTTCAAGAACTGAAGACCGCAAGTACGCCAAGATGGC 1740
Qy 1741 ACCGCCACACCAAGCAGTGAAGCAGTGAACCGGCGGCTGCAGAAATGCCCATGGAG 1800
Db 1741 ACCGCCACACCAAGCAGTGAAGCAGTGAACCGGCGGCTGCAGAAATGCCCATGGAG 1800
Qy 1801 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAGGAGACCTGG 1860
Db 1801 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAGGAGACCTGG 1860
Qy 1861 GAGACTGTGGACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1920
Db 1861 GAGACTGTGGACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAAC 1920
Qy 1921 ACCCCCCCTGTGTGAGCTGTGTACCGCTGAGAGGAGGCCCATCATCGGCGCCGAG 1980
Db 1921 ACCCCCCCTGTGTGAGCTGTGTACCGCTGAGAGGAGGCCCATCATCGGCGCCGAG 1980
Qy 1981 ACCTTCTACGTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTG 2040
Db 1981 ACCTTCTACGTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTG 2040
Qy 2041 ACCGACCGGGCCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACCGGGCCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Qy 2101 CTGCAGGCCATCCAGCTGGCCCTTCAGGACAGCGGCGAGGCTGACATCGTGACCGAC 2160
Db 2101 CTGCAGGCCATCCAGCTGGCCCTTCAGGACAGCGGCGAGGCTGACATCGTGACCGAC 2160
Qy 2161 AGCCAGTACGCCCTGGGCGCATCATCAGGCCAGCCCGACAGAGCGAGCGAGCTGGTG 2220
Db 2161 AGCCAGTACGCCCTGGGCGCATCATCAGGCCAGCCCGACAGAGCGAGCGAGCTGGTG 2220
Qy 2221 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGGTGCCCGCC 2280
Db 2221 AACAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGGTGCCCGCC 2280
Qy 2281 CACAAGGCGCATCGCGCGACAGCAGATCGACAAGCTGTGTGAGCAAGGGGATCCGCAAG 2340
Db 2281 CACAAGGCGCATCGCGCGACAGCAGATCGACAAGCTGTGTGAGCAAGGGGATCCGCAAG 2340
Qy 2341 GTGCTGTTCTTGGACGCGCATCGATGGCGGCTGATCTACCAAGTACATGAGACGACCTG 2400
Db 2341 GTGCTGTTCTTGGACGCGCATCGATGGCGGCTGATCTACCAAGTACATGAGACGACCTG 2400
Qy 2401 TACGTGGCGAGCGCGCCCTTAGGATCGATTTAAAGCTTCCCGGGCTAGCACCGGT 2457
Db 2401 TACGTGGCGAGCGCGCCCTTAGGATCGATTTAAAGCTTCCCGGGCTAGCACCGGT 2457
```

## RESULT 5

```
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
XX 15-MAY-2002 (first entry)
XX
XX Synthetic construct PR975(+) SEQ ID NO:30.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gsg; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
XX Human immunodeficiency virus; type C.
OS
XX Synthetic.
XX
XX WO200204493-A2.
```



PD	17-JAN-2002.	QY	541	GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCATCGGACCGTG	600
XX		DB	541	GTGCGCAGTACGACCAAGATCTCTGATCGAGATCTGCGGCAAGAGGCATCGGACCGTG	600
PF	05-JUL-2001; 2001WO-US021241.				
XX		QY	601	CTGATCGGCCCCACCCCGTGAAATCATCTGCGCGCAACATCTGCTACCCAGCTGGGTGC	660
XX		DB	601	CTGATCGGCCCCACCCCGTGAAATCATCTGCGCGCAACATCTGCTACCCAGCTGGGTGC	660
PA	(CHIR ) CHIRON CORP.				
PA	(UYST-) UNIV STELLENBOSCH.				
XX		QY	661	ACCTGAACTTCCCATCATCGAGACCGTCCCTGAGCTGAAGCTGAAGCCCGGATG	720
PI	Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	DB	661	ACCTGAACTTCCCATCATCGAGACCGTCCCTGAGCTGAAGCTGAAGCCCGGATG	720
XX	WPI; 2002-154920/20.				
XX		QY	721	GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACGAGGAGAGATCAAGGCCCTGACCGCC	780
XX		DB	721	GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACGAGGAGAGATCAAGGCCCTGACCGCC	780
PT	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in				
PT	applications including DNA immunization or generation of packaging cell				
PT	lines, particularly in gene therapy.				
XX		QY	781	ATCTGCGAGGAGATGAGAGAGGCGCAAGATCAACAGATCGGCCCCCGAGAACCCCTAC	840
XX	Claim 1; Fig 8; 233pp; English.	DB	781	ATCTGCGAGGAGATGAGAGAGGCGCAAGATCAACAGATCGGCCCCCGAGAACCCCTAC	840
CC	The present invention describes expression cassettes comprising a				
CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	QY	841	AACACCCCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900
CC	type C polypeptides. The expression cassettes comprise any of the HIV	DB	841	AACACCCCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	900
CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef				
CC	(i). (i) have immunostimulant activity and can be used in gene therapy.	QY	901	TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCGATCCCCAC	960
CC	The HIV type C polynucleotides are useful in applications including DNA	DB	901	TTCCGCGAGCTGAACCAAGCGCACCCAGGACTTCTGCGAGGTGAGCTGGGCGATCCCCAC	960
CC	immunisation, generation of packaging cell lines, and production of HIV				
CC	Type C proteins. The polynucleotides are particularly useful in gene	QY	961	CCCGCGGCTGTAAG	1020
CC	therapy and DNA immunisation applications. AB19942 to ABL40054 and	DB	961	CCCGCGGCTGTAAG	1020
CC	ABB06204 to ABB06215 represent sequences used in the exemplification of				
CC	the present invention	QY	1021	AGGTGCTGCTGAG	1080
XX	Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;	DB	1021	AGGTGCTGCTGAG	1080
Query Match	99.2%; Score 2442.2; DB 6; Length 2469;				
Best Local Similarity	99.6%; Pred. No. 1.5e-294;				
Matches 2460; Conservative	0; Mismatches 3; Indels 6; Gaps 1;				
QY	1	QY	1	GTGACGCCCACTGCGCGAGGCGCATGAGCAGGCCAGCCAGCGGCAACATCTCTGATCGAG	60
DB	1	DB	1	GTGACGCCCACTGCGCGAGGCGCATGAGCAGGCCAGCCAGCGGCAACATCTCTGATCGAG	60
QY	61	QY	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGCAAGAGGGC	120
DB	61	DB	61	CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTCGCGCAAGAGGGC	120
QY	121	QY	121	CACATCGCGCGCAACTGCGCGGCCCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	180
DB	121	DB	121	CACATCGCGCGCAACTGCGCGGCCCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	180
QY	181	QY	181	GGCCACAGATGAGAGAGTGCACCGAGCGCCAGGCAACTTCTTCGCGAGGAGCTGGCC	240
DB	181	DB	181	GGCCACAGATGAGAGAGTGCACCGAGCGCCAGGCAACTTCTTCGCGAGGAGCTGGCC	240
QY	241	QY	241	TTCCCGCAGGAGGCGCGAGTTCGCCAGCAGCAGAACCGCGCAACAGCCCCCACC	300
DB	241	DB	241	TTCCCGCAGGAGGCGCGAGTTCGCCAGCAGCAGAACCGCGCAACAGCCCCCACC	300
QY	301	QY	301	AGCCCGAGTGCAGGTGCGCGCGCAACACCCCGCAGCAGAGCGCGCGCGCGAGCGCAG	360
DB	301	DB	301	AGCCCGAGTGCAGGTGCGCGCGCAACACCCCGCAGCAGAGCGCGCGCGAGCGCAG	360
QY	361	QY	361	GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG	420
DB	361	DB	361	GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG	420
QY	421	QY	421	GGCGGCGAGATCAAGGAGGCGCTCTGTGACACCGCGCGCAGACACCGTGTGAGGAG	480
DB	421	DB	421	GGCGGCGAGATCAAGGAGGCGCTCTGTGACACCGCGCGCAGACACCGTGTGAGGAG	480
QY	481	QY	481	ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG	540
DB	481	DB	481	ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAG	540

```
Db 1621 GTGTTACTAGACCCAGCAGAGACCTGGTGGCCGAGATCCAGAACGACGGCCACGACCAG 1680
Qy 1675 TGGACCTTACAGATCTACACGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1734
Db 1681 TGGACCTTACAGATCTACACGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAG 1740
Qy 1735 ATGGCAGCGCCGACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCC 1794
Db 1741 ATGGCAGCGCCGACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCC 1800
Qy 1795 ATGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1854
Db 1801 ATGAGAGCATCTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAG 1860
Qy 1855 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1914
Db 1861 ACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTC 1920
Qy 1915 GTGAACACCCCCCTCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC 1974
Db 1921 GTGAACACCCCCCTCTGGTGAAGCTGTGGTACAGCTGGAGAGGAGCCCATCATCGGC 1980
Qy 1975 GCCGAGACCTTCTAGCTGAGCGCGCGCCGACCGCAGAGCCAGATCGGCAAGGCCGCG 2034
Db 1981 GCCGAGACCTTCTAGCTGAGCGCGCGCCGACCGCAGAGCCAGATCGGCAAGGCCGCG 2040
Qy 2035 TACGTGACCGACCGGGCGGCGAGAAAGATCTGAGCCTGACCGAGACCAACCAAGAG 2094
Db 2041 TACGTGACCGACCGGGCGGCGAGAAAGATCTGAGCCTGACCGAGACCAACCAAGAG 2100
Qy 2095 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGAGGACAGCGGAGCGAGGTGAACATCGTG 2154
Db 2101 ACCGAGCTGACGAGCCATCCAGCTGGCCCTGAGGACAGCGGAGCGAGGTGAACATCGTG 2160
Qy 2155 ACCGACGACGAGTACGCTGGCCATCATCAGGCCCGACCGCAGAGCGAGCGAG 2214
Db 2161 ACCGACGACGAGTACGCTGGCCATCATCAGGCCCGACCGCAGAGCGAGCGAG 2220
Qy 2215 CTGCTGAACACGAGATCATCAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2274
Db 2221 CTGCTGAACACGAGATCATCAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTG 2280
Qy 2275 CCGCGCCCAAGAGGATCGCGGCAACGAGCAGATCGAAGCTGTGTAGCAAGGGCATC 2334
Db 2281 CCGCGCCCAAGAGGATCGCGGCAACGAGCAGATCGAAGCTGTGTAGCAAGGGCATC 2340
Qy 2335 CGCAAGTGTCTTCTGAGCGGATCGATGGCGGCATCGTATACAGTACAGTACATGAC 2394
Db 2341 CGCAAGTGTCTTCTGAGCGGATCGATGGCGGCATCGTATACAGTACATGAC 2400
Qy 2395 GACCTGTACGTGGGCGGCGGCGCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2454
Db 2401 GACCTGTACGTGGGCGGCGGCGCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Qy 2455 GGTGAATTC 2463
Db 2461 GGTGAATTC 2469
```

## RESULT 6

ADM73764

ID ADM73764 standard; DNA; 2469 BP.

XX AC

XX AC ADM73764;

XX DT

XX 03-JUN-2004 (first entry)

XX DE

XX HIV-1 polynucleotide #7.

XX KW

XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;

XX KW

XX HIV type C protein; immunostimulant.

XX XX

OS Human immunodeficiency virus 1.

XX PN

XX US2003223961-A1.

XX PD

XX 04-DEC-2003.

XX PF

XX 05-JUL-2001; 2001US-00899575.

XX PR

XX 05-JUL-2000; 2000US-00610313.

XX PA

XX (MEGE/) MEGEDE J Z.

XX PA

XX (BARN/) BARNETT S W.

XX PA

XX (ENGE/) ENGELBRECHT S.

XX XX

XX (RENS/) RENSBERG E J V.

PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJ;

XX WPI; 2004-060515/06.

XX DR

XX WPI; 2004-060515/06.

XX PT

XX New expression cassette comprising a polynucleotide sequence encoding an

XX PT

XX HIV Pol polypeptide, useful in eliciting an immune response, in DNA

XX PT

XX immunization, generating of packaging cell lines or in producing HIV Type

XX PT

XX C proteins.

XX PS

XX Claim 1; SEQ ID NO 30; 160pp; English.

XX CC

XX The invention relates to an expression cassette comprising a

XX CC

XX polynucleotide sequence encoding an HIV Pol polypeptide. The invention

XX CC

XX also relates to a recombinant expression system for use in a host cell

XX CC

XX comprising an expression cassette, where the polynucleotide sequence

XX CC

XX further comprises control elements capable of driving expression in the

XX CC

XX selected host cell, a cell comprising an expression cassette where the

XX CC

XX polynucleotide sequence further comprises control elements compatible

XX CC

XX with the expression in the cell and a composition for generating an

XX CC

XX immunological response, comprising an expression cassette. The expression

XX CC

XX cassette and the methods of the invention are useful in eliciting an

XX CC

XX immune response, in DNA immunisation, in generation of packaging cell

XX CC

XX lines and in producing HIV Type C proteins. This sequence represents an

XX CC

XX HIV-1 polynucleotide of the invention.

XX SQ

XX Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 99.2%; Score 2442.2; DB 12; Length 2469;

Best Local Similarity 99.6%; Pred. No. 1.5e-294;

Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTCCAGCCACCATGGCGGCGCATGAGCGCCATGAGCGCCACCGCGCAACATCTGATGCAG 60

Db 1 GTCAGCCACCATGGCGGCGCATGAGCGCCATGAGCGCCACCGCGCAACATCTGATGCAG 60

Qy 61 CGCAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120

Db 61 CGCAGCAACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120

Qy 121 CACATCGCCGCAACTGCGGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180

Db 121 CACATCGCCGCAACTGCGGCGCCCCCGCAAGAGGGCTCTGGAAGTGGCGCAAGGAG 180

Qy 181 GGCCACCATGTAAGGACTGACCGAGCGCCAGCGCAACTTCTTCCGCGAGGACTGSCC 240

Db 181 GGCCACCATGTAAGGACTGACCGAGCGCCAGCGCAACTTCTTCCGCGAGGACTGSCC 240

Qy 241 TTCCCGCCAGGCAAGGCGCCGAGTTCCTCCAGCAGCAGAACCGCGCCCAAGCCCCACC 300

Db 241 TTCCCGCCAGGCAAGGCGCCGAGTTCCTCCAGCAGCAGAACCGCGCCCAAGCCCCACC 300

Qy 301 AGCCGCGAGTGTGAGGTGCGCGGCAAAACCCCGCAGCGAGGCGCGCGCGAGGCGCAG 360

Db 301 AGCCGCGAGTGTGAGGTGCGCGGCAAAACCCCGCAGCGAGGCGCGCGCGAGGCGCAG 360

Qy 361 GGCAACCTGACCTTCCCGCAGATCACCTGTGGCGCGCCCTCTGTGAGCATCAAGGTG 420

Db 361 GGCAACCTGACCTTCCCGCAGATCACCTGTGGCGCGCCCTCTGTGAGCATCAAGGTG 420

Qy	421	GGCGCCAGATCAAGAGGAGCCCTGCTGGACACCGGCGCCGACGACACCGTGCCTGGAGGAG	480
Db	421	GGCGCCAGATCAAGAGGAGCCCTGCTGGACACCGGCGCCGACGACACCGTGCCTGGAGGAG	480
Qy	481	ATGAGCCTGCCCGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG	540
Db	481	ATGAGCCTGCCCGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCAAG	540
Qy	541	GTGCGCAGTACGACACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG	600
Db	541	GTGCGCAGTACGACACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG	600
Qy	601	CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660
Qy	661	ACCTGAATCTCCCATCAGCCCATCGAGACGTCGCGCGTGCCTGAGCTGAGAGCCGCGCATG	720
Db	661	ACCTGAATCTCCCATCAGCCCATCGAGACGTCGCGCGTGCCTGAGCTGAGAGCCGCGCATG	720
Qy	721	GACGGCCCCAAGGTGAAGCAAGTGCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCC	780
Db	721	GACGGCCCCAAGGTGAAGCAAGTGCCCTGACCCGAGGAGAGATCAAGGCCCTGACCGCC	780
Qy	781	ATCTCGAGGAGATGGAAGAGAGGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTAC	840
Db	781	ATCTCGAGGAGATGGAAGAGAGGGCAAGATCAACCAAGATCGGCCCCCGAGAACCCCTAC	840
Qy	841	AACACCCCGTGTTCGCCCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Db	841	AACACCCCGTGTTCGCCCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGTGGAC	900
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGCACTTCTGGGAGGTGCAAGCTGGGCATCCCCCCAC	960
Db	901	TTCCGCGAGCTGAACAAGCGCACCCAGCACTTCTGGGAGGTGCAAGCTGGGCATCCCCCCAC	960
Qy	961	CCCGCGCGCTGAAGAAAGAGAGCGTGACCGTGTGGAAGTGGGCGAGCGCTACTTC	1020
Db	961	CCCGCGCGCTGAAGAAAGAGAGCGTGACCGTGTGGAAGTGGGCGAGCGCTACTTC	1020
Qy	1021	AGCTGCGCTTGGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCCAGCATCAAC	1080
Db	1021	AGCTGCGCTTGGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCCAGCATCAAC	1080
Qy	1081	AACGAGACCCCGGCATCCGCTACAGTACAAAGTGTGCCCCCAGGGCTGGAAGGGCAGC	1140
Db	1081	AACGAGACCCCGGCATCCGCTACAGTACAAAGTGTGCCCCCAGGGCTGGAAGGGCAGC	1140
Qy	1141	CCGAGCATCTTCGAGAGGAGATGACCAAGATCTCTGAGCGCTTCGCGGCCCGCAACCCC	1200
Db	1141	CCGAGCATCTTCGAGAGGAGATGACCAAGATCTCTGAGCGCTTCGCGGCCCGCAACCCC	1200
Qy	1201	GAGATCTGTACTACCA-----GGCCCCCTGTACGTGGGCGAGCGACCTGGAGATCGCG	1254
Db	1201	GAGATCTGTACTACCAAGTACATGACGACTGTACGTGGGCGAGCGACCTGGAGATCGCG	1254
Qy	1255	CAGCACCGCGCAAGATCGAGGAGCTCGCAAGCACTGTGTGGCTGGGCTTCACCAACC	1314
Db	1261	CAGCACCGCGCAAGATCGAGGAGCTCGCAAGCACTGTGTGGCTGGGCTTCACCAACC	1320
Qy	1315	CCGACAAGAGCACCAAGAGGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGCACCCC	1374
Db	1321	CCGACAAGAGCACCAAGAGGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGCACCCC	1380
Qy	1375	GACAAGTGGACCGGTGACGCCCATCGAGCTGCCCCGAGAAGGAGAGCTGGAACCGTGAACGAC	1434
Db	1381	GACAAGTGGACCGGTGACGCCCATCGAGCTGCCCGAGAAGGAGAGCTGGAACCGTGAACGAC	1440
Qy	1435	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCCGGCATCAAGGTG	1494
Db	1441	ATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCCGGCATCAAGGTG	1500

QY	1495	CGCCAGCTGTGCAAGCTGTCTGCGGGCGCCAAAGCCCTTGACCGACATCTGTCGCCCTTGACC	1555
DB	1501	CGCCAGCTGTGCAAGCTGTCTGCGGGCGCCAAAGCCCTTGACCGACATCTGTCGCCCTTGACC	1560
QY	1555	GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTTCGCGAGGCCGTGACCGC	1614
DB	1561	GAGGAGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTTCGCGAGGCCGTGACCGC	1620
QY	1615	GTGTACTACGACCCGACCAAGGACTCTGGTGGCCGAGATCCAGAGCAGGCCCCAGCCAG	1674
DB	1621	GTGTACTACGACCCGACCAAGGACTCTGGTGGCCGAGATCCAGAGCAGGCCCCAGCCAG	1680
QY	1675	TGGACTTACGAGTCTACGAGGACCTTCAAAGAACCTGAAAGCCGCAAGTAGGCCCCAAG	1734
DB	1681	TGGACTTACGAGTCTACGAGGACCTTCAAAGAACCTGAAAGCCGCAAGTAGGCCCCAAG	1740
QY	1735	ATGGGCACCGCCACACACCAACGACGCTGAAGCAGCTGACCGAGGCGGTGACAGAAAGATCGCC	1794
DB	1741	ATGGGCACCGCCACACCAACGACGCTGAAGCAGCTGACCGAGGCGGTGACAGAAAGATCGCC	1800
QY	1795	ATGAGAGCATCTGATCTCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAG	1854
DB	1801	ATGAGAGCATCTGATCTCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAG	1860
QY	1855	ACCTGGGAGACCTTGGTGGACCGACTCTGCGAGCCACCTTGGATCCCCGAGTGGAGTTC	1914
DB	1861	ACCTGGGAGACCTTGGTGGACCGACTCTGCGAGCCACCTTGGATCCCCGAGTGGAGTTC	1920
QY	1915	GTGAACACCCCCCTTGGTGAAGCTGTGTATCCAGCTGGAGAGGAGCCCATCATCGGC	1974
DB	1921	GTGAACACCCCCCTTGGTGAAGCTGTGTATCCAGCTGGAGAGGAGCCCATCATCGGC	1980
QY	1975	GCCGAGACCTTTACGTGGAAGCGGCGCCGCAACCGCGAGACCAAGATCGCGAGGCCCGC	2034
DB	1981	GCCGAGACCTTTACGTGGAAGCGGCGCCGCAACCGCGAGACCAAGATCGCGAGGCCCGC	2040
QY	2035	TACGTGACCGACCGGGCCGCGAGAGATCGTGAGCCTGACCGAGACCAACCAACGAGAG	2094
DB	2041	TACGTGACCGACCGGGCCGCGAGAGATCGTGAGCCTGACCGAGACCAACCAACGAGAG	2100
QY	2095	ACCGAGCTGCAGGCCATTCAGCTGGCCCTCGAGSACAGCGGCGAGCGGTGAACATCGTG	2154
DB	2101	ACCGAGCTGCAGGCCATTCAGCTGGCCCTCGAGSACAGCGGCGAGCGGTGAACATCGTG	2160
QY	2155	ACCGAGCCAGTACGCCCTTGGGCATCATCCAGGCCAGCCGACAGAGGCGAGAGCGAG	2214
DB	2161	ACCGAGCCAGTACGCCCTTGGGCATCATCCAGGCCAGCCGACAGAGGCGAGAGCGAG	2220
QY	2215	CTGTGTGAACAGATCATCGAGCAGCTCATCAAGAAAGGAGAGGTGTACCTGAGCTGGGTG	2274
DB	2221	CTGTGTGAACAGATCATCGAGCAGCTCATCAAGAAAGGAGAGGTGTACCTGAGCTGGGTG	2280
QY	2275	CCCGCCCAAGGGGCATCGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGCATC	2334
DB	2281	CCCGCCCAAGGGGCATCGCGGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGCATC	2340
QY	2335	CGCAAGCTGTTCCTGACCGCATCGATGGGGCATCGTGTATCTACCAAGTACATCGAC	2394
DB	2341	CGCAAGCTGTTCCTGACCGCATCGATGGGGCATCGTGTATCTACCAAGTACATCGAC	2400
QY	2395	GACCTGTACGTGGGACGGCGGCCCTTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACC	2454
DB	2401	GACCTGTACGTGGGACGGCGGCCCTTAGGATCGATTTAAAGCTTCCCGGGGCTAGCACC	2460
QY	2455	GGTGAATTC	2463
DB	2461	GGTGAATTC	2469

RESULT 7  
ABL39961  
ID ABL39961 standard; DNA; 2457 BP.  
XX



```
Db 1375 TGGACCGTCGAGCCATCGAGTGTCCCGAGAGAGAGCTGGACCGTGAACGACATCCAG 1434
Qy 1441 AAGCTGTGGCGAGCTGAAGTGTGGCCAGCAGATCTACCCCGCATCAAGGTGCGCCAG 1500
Db 1435 AAGCTGTGGCGAGCTGAAGTGTGGCCAGCAGATCTACCCCGCATCAAGGTGCGCCAG 1494
Qy 1501 CTGTGAAGCTGTCTGCGCGCGCCAGGCCCTGACCGACATCTGTGCCCTTGACCGAGGAG 1560
Db 1495 CTGTGAAGCTGTCTGCGCGCGCCAGGCCCTGACCGACATCTGTGCCCTTGACCGAGGAG 1554
Qy 1561 GCCAGCTGAGCTGTGGCCGAGAACCCGAGATCTGTGGAGCCCGTGCACGCGGTGTAC 1620
Db 1555 GCCAGCTGAGCTGTGGCCGAGAACCCGAGATCTGTGGAGCCCGTGCACGCGGTGTAC 1614
Qy 1621 TAGACCCCGCAGGAGCTGTGGCGGAGATCCAGAGCAGGCGCACGACGAGTGAC 1680
Db 1615 TAGACCCCGCAGGAGCTGTGGCGGAGATCCAGAGCAGGCGCACGACGAGTGAC 1674
Qy 1681 TACCAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGC 1740
Db 1675 TACCAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGC 1734
Qy 1741 ACCGCCACACCAACGAGCTGAAGCAGTGAACGAGCCGTCGCAAGATGCGCAATGAG 1800
Db 1735 ACCGCCACACCAACGAGCTGAAGCAGTGAACGAGCCGTCGCAAGATGCGCAATGAG 1794
Qy 1801 AGCATCTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1860
Db 1795 AGCATCTGTATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGG 1854
Qy 1861 GAGACCTGTGTGAGCCGACTACTGTGAGCGCACCTGTGATCCCGAGTGGGAGTTCGTGAAC 1920
Db 1855 GAGACCTGTGTGAGCCGACTACTGTGAGCGCACCTGTGATCCCGAGTGGGAGTTCGTGAAC 1914
Qy 1921 ACCCCCCCTGTGTGAGCTGTGTGACAGCTGAGAGAGAGCCCATCATCGGCGCCGAG 1980
Db 1915 ACCCCCCCTGTGTGAGCTGTGTGACAGCTGAGAGAGAGCCCATCATCGGCGCCGAG 1974
Qy 1981 ACCTTCTACGTGAGCGCGCCGCAACCGGAGACCAAGATCGGCAAGCGCGGTACGTG 2040
Db 1975 ACCTTCTACGTGAGCGCGCGCCGCAACCGGAGACCAAGATCGGCAAGCGCGGTACGTG 2034
Qy 2041 ACCGACCGGCGCGCGCAGAGATCGTGAGCTGACCGAGACCAACACCAAGAGACCGAG 2100
Db 2035 ACCGACCGGCGCGCGCAGAGATCGTGAGCTGACCGAGACCAACACCAAGAGACCGAG 2094
Qy 2101 CTGAGGCCATTCAGCTGGCCCTGCAGACAGCGGACAGGAGTGAACATCGTGACCGAC 2160
Db 2095 CTGAGGCCATTCAGCTGGCCCTGCAGACAGCGGACAGGAGTGAACATCGTGACCGAC 2154
Qy 2161 AGCNGTACGCTGGGCATCATCCAGGCCAGCCGACAGAGAGGAGGAGCTGGTG 2220
Db 2155 AGCNGTACGCTGGGCATCATCCAGGCCAGCCGACAGAGGAGGAGCTGGTG 2214
Qy 2221 AACCAGATCATCGAGCAGCTGATCAAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2280
Db 2215 AACCAGATCATCGAGCAGCTGATCAAGAGAGAGTGTACCTGAGCTGGGTGCCCGCC 2274
Qy 2281 CACAAGGGCATTCGCGCGCAACGAGCAGATCGACAAGCTGTGTAGCAAGGGGATCCGCAAG 2340
Db 2275 CACAAGGGCATTCGCGCGCAACGAGCAGATCGACAAGCTGTGTAGCAAGGGGATCCGCAAG 2334
Qy 2341 GTGCTGTCTTGGAGCGCATCGATGGCGGATCGTGATCTACCAAGTACATGGAGCAGCTG 2400
Db 2335 GTGCTGTCTTGGAGCGCATCGATGGCGGATCGTGATCTACCAAGTACATGGAGCAGCTG 2394
Qy 2401 TACGTGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460
Db 2395 TACGTGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2454
Qy 2461 TTC 2463
Db 2455 TTC 2457
```

RESULT 8  
ADM73766

ID ADM73766 standard; DNA; 2457 BP.

XX ADM73766;

XX AC ADM73766;

DT 03-JUN-2004 (first entry)

XX HIV-1 polynucleotide #9.

XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;

KW HIV type C protein; immunostimulant.

XX Human immunodeficiency virus 1.

OS US200323961-A1.

XX 04-DEC-2003.

XX 05-JUL-2001; 2001US-00899575.

XX 05-JUL-2000; 2000US-00610313.

XX (MEGE/) MEGEDE J Z.

PA (BARN/) BARNETT S W.

PA (ENGE/) ENGELBRECHT S.

PA (RENS/) RENSBURG E J V.

XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJ;

XX WPI; 2004-060515/06.

XX New expression cassette comprising a polynucleotide sequence encoding an HIV Pol polypeptide, useful in eliciting an immune response, in DNA immunization, generating of packaging cell lines or in producing HIV Type C proteins.

PS Claim 1; SEQ ID NO 32; 160pp; English.

XX The invention relates to an expression cassette comprising a

CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention

CC also relates to a recombinant expression system for use in a host cell

CC comprising an expression cassette, where the polynucleotide sequence

CC further comprises control elements capable of driving expression in the

CC selected host cell, a cell comprising an expression cassette where the

CC polynucleotide sequence further comprises control elements compatible

CC with the expression in the cell and a composition for generating an

CC immunological response, comprising an expression cassette. The expression

CC cassette and the methods of the invention are useful in eliciting an

CC immune response, in DNA immunisation, in generation of packaging cell

CC lines and in producing HIV Type C proteins. This sequence represents an

XX HIV-1 polynucleotide of the invention.

SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 98.9%; Score 2436.2; DB 12; Length 2457;

Best Local Similarity 99.6%; Pred. No. 8.3e-294;

Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

Qy 1 GTGACGCCCACTGGCGGAGGCCATGAGCAGGCCACCGGCCCAACATCTCTGATCGAG 60

Db 1 GTGACGCCCACTGGCGGAGGCCATGAGCAGGCCACCGGCCCAACATCTCTGATCGAG 60

Qy 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGGC 120

Db 61 CGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGGC 120

Qy 121 CACATCGCCCACTCGCGGCGCCCCCGCAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180

Db 121 CACATCGCCCACTCGCGGCGCCCCCGCAGAGGGCTGCTGGAAGTGGCGCAAGGAG 180

QY 181 GGCACACGATGAGGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240  
 Db 181 GGCACACGATGAGGACTGACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCC 240  
 QY 241 TTCCCCAGGGCAAGCGCCGAGTTCCACGAGAGGAGAACCGCGCCAAAGAGCCCAACC 300  
 Db 241 TTCCCCAGGGCAAGCGCCGAGTTCCACGAGAGGAGAACCGCGCCAAAGAGCCCAACC 300  
 QY 301 AGCCGCGAGCTGACGGTGGGGGCGAACAACCCCGCAGCGAGCGCGCGCGAGGCCAG 360  
 Db 301 AGCCGCGAGCTGACGGTGGGGGCGAACAACCCCGCAGCGAGCGCGCGAGGCCAG 360  
 QY 361 GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTTGTGAGCATCAAGTG 420  
 Db 361 GGCACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTTGTGAGCATCAAGTG 420  
 QY 421 GCGCGCAGATCAAGAGCGCCCTGTGACACCGCGCGCGAGCACACCTGTGAGGAG 480  
 Db 421 GCGCGCAGATCAAGAGCGCCCTGTGACACCGCGCGCGAGCACACCTGTGAGGAG 480  
 QY 481 ATGAGCTTCCCGGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGCTTCAACAAG 540  
 Db 481 ATGAGCTTCCCGGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGCTTCAACAAG 540  
 QY 541 GTGCGCAGTACGACAGATCTGTATGAGATCTCGCGGAGAGAGGCCATCGGCAAGTG 600  
 Db 541 GTGCGCAGTACGACAGATCTGTATGAGATCTCGCGGAGAGAGGCCATCGGCAAGTG 600  
 QY 601 CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATCTGACACCACTGGGCTGC 660  
 Db 601 CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATCTGACACCACTGGGCTGC 660  
 QY 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
 Db 661 ACCCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATG 720  
 QY 721 GACGCCCCAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 Db 721 GACGCCCCAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
 QY 781 ATCTGCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAACCCCTAC 840  
 Db 781 ATCTGCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGAGAACCCCTAC 840  
 QY 841 AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGAC 900  
 Db 841 AACACCCCGTGTTCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGAC 900  
 QY 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCGAGCTGGGCAATCCCCAC 960  
 Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTGCGAGCTGGGCAATCCCCAC 960  
 QY 961 CCGCGCGCTGAGAGAGAGAGCTGACCGTGTGAGCGCTGGGCGAGCGCTACTTC 1020  
 Db 961 CCGCGCGCTGAGAGAGAGAGCTGACCGTGTGAGCGCTGGGCGAGCGCTACTTC 1020  
 QY 1021 AGCGTCCCTTGGAGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
 Db 1021 AGCGTCCCTTGGAGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
 QY 1081 AACGAGACCCCGGATCCGTTACAGTACAAAGTGTGCCCGAGGCTGGAAGGGGAGC 1140  
 Db 1081 AACGAGACCCCGGATCCGTTACAGTACAAAGTGTGCCCGAGGCTGGAAGGGGAGC 1140  
 QY 1141 CCCAGCATCTCCAGAGCAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCAACCC 1200  
 Db 1141 CCCAGCATCTCCAGAGCAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCAACCC 1200  
 QY 1201 GAGATCGTGTATCTACACGCGCCCTGTGACGTTGGGAGCGACCTGGAGATCGGCCAGC 1260  
 Db 1201 GAGATCGTGTATCTACACGCGCCCTGTGACGTTGGGAGCGACCTGGAGATCGGCCAGC 1260  
 QY 1261 GCGCCCAAGATCGAGGAGCTGGCAAGCACCTGTGCGCTGGGGCTTCCACACCCCGAC 1320

Db 1261 GCGCCCAAGATCGAGGAGCTGGCAAGCACCTGTGCGCTGGGGCTTCCACACCCCGAC 1320  
 QY 1321 AAGAAGACACCAAGAGAGCCCTTCTGTGTGGTGGGTACGAGCTGACCCCGACCAAG 1380  
 Db 1321 AAGAAGACACCAAGAGAGCCCTTCTGTGTGGTGGGTACGAGCTGACCCCGACCAAG 1380  
 QY 1374 AAGAAGACACCAAGAGAGCCCTTCTGTGTGGTGGGTACGAGCTGACCCCGACCAAG 1374  
 Db 1374 AAGAAGACACCAAGAGAGCCCTTCTGTGTGGTGGGTACGAGCTGACCCCGACCAAG 1374  
 QY 1381 TGGACCGTGGAGCCCATCGAGCTGCGCCGAGAGGAGAGCTGGAACGAGATCCAG 1440  
 Db 1381 TGGACCGTGGAGCCCATCGAGCTGCGCCGAGAGGAGAGCTGGAACGAGATCCAG 1440  
 QY 1441 AAGCTGTGGGCAAGCTGAACTGGGCGAGCGAGATCTACCCCGGATCAAGGTGGCCAG 1500  
 Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCGAGCGAGATCTACCCCGGATCAAGGTGGCCAG 1500  
 QY 1435 AAGCTGTGGGCAAGCTGAACTGGGCGAGCGAGATCTACCCCGGATCAAGGTGGCCAG 1494  
 Db 1435 AAGCTGTGGGCAAGCTGAACTGGGCGAGCGAGATCTACCCCGGATCAAGGTGGCCAG 1494  
 QY 1501 CTGTGCAAGCTGCTGCGCGCGCCAAAGGCCCTGACCCGATCTGTCCTGACCCGAGAG 1560  
 Db 1501 CTGTGCAAGCTGCTGCGCGCGCCAAAGGCCCTGACCCGATCTGTCCTGACCCGAGAG 1560  
 QY 1495 CTGTGCAAGCTGCTGCGCGCGCCAAAGGCCCTGACCCGATCTGTCCTGACCCGAGAG 1554  
 Db 1495 CTGTGCAAGCTGCTGCGCGCGCCAAAGGCCCTGACCCGATCTGTCCTGACCCGAGAG 1554  
 QY 1561 GCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGAACGAGCTGTAC 1620  
 Db 1561 GCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGAACGAGCTGTAC 1620  
 QY 1555 GCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGAACGAGCTGTAC 1614  
 Db 1555 GCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGAACGAGCTGTAC 1614  
 QY 1621 TAGCACCCAGCAAGGACCTGCTGGCGAGATCCAGAGAGCGGSCACGACCAAGTGGACC 1680  
 Db 1621 TAGCACCCAGCAAGGACCTGCTGGCGAGATCCAGAGAGCGGSCACGACCAAGTGGACC 1680  
 QY 1615 TAGCACCCAGCAAGGACCTGCTGGCGAGATCCAGAGAGCGGSCACGACCAAGTGGACC 1674  
 Db 1615 TAGCACCCAGCAAGGACCTGCTGGCGAGATCCAGAGAGCGGSCACGACCAAGTGGACC 1674  
 QY 1681 TACCAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGAGCGGCAAGTACGCAAGATGGCC 1740  
 Db 1681 TACCAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGAGCGGCAAGTACGCAAGATGGCC 1740  
 QY 1675 TACCAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGAGCGGCAAGTACGCAAGATGGCC 1734  
 Db 1675 TACCAGATCTACCAAGGAGCCCTTCAAGAACCTGAAGAGCGGCAAGTACGCAAGATGGCC 1734  
 QY 1741 ACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCGCTGCAAGAGATCGCCATGGAG 1800  
 Db 1741 ACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCGCTGCAAGAGATCGCCATGGAG 1800  
 QY 1735 ACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCGCTGCAAGAGATCGCCATGGAG 1794  
 Db 1735 ACCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGCGCTGCAAGAGATCGCCATGGAG 1794  
 QY 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCTCCAGAGAGAGACTGG 1860  
 Db 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCTCCAGAGAGAGACTGG 1860  
 QY 1795 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCTCCAGAGAGAGACTGG 1854  
 Db 1795 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCTCCAGAGAGAGACTGG 1854  
 QY 1861 GAGACCTGTGGAGCCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1920  
 Db 1861 GAGACCTGTGGAGCCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1920  
 QY 1855 GAGACCTGTGGAGCCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1914  
 Db 1855 GAGACCTGTGGAGCCGACTACTGCGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1914  
 QY 1921 ACCGCCCCCTTGGTGAAGCTGTGTAACGCTGGAGAGAGGAGCCCATCATCGGCCCGAG 1980  
 Db 1921 ACCGCCCCCTTGGTGAAGCTGTGTAACGCTGGAGAGAGGAGCCCATCATCGGCCCGAG 1980  
 QY 1915 ACCGCCCCCTTGGTGAAGCTGTGTAACGCTGGAGAGAGGAGCCCATCATCGGCCCGAG 1974  
 Db 1915 ACCGCCCCCTTGGTGAAGCTGTGTAACGCTGGAGAGAGGAGCCCATCATCGGCCCGAG 1974  
 QY 1981 ACCTTCTACGTGGAGCGGGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGCTG 2040  
 Db 1981 ACCTTCTACGTGGAGCGGGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGCTG 2040  
 QY 1975 ACCTTCTACGTGGAGCGGGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGCTG 2034  
 Db 1975 ACCTTCTACGTGGAGCGGGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGCTG 2034  
 QY 2041 ACCGACCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCCAG 2100  
 Db 2041 ACCGACCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCCAG 2100  
 QY 2035 ACCGACCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCCAG 2094  
 Db 2035 ACCGACCGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCCAG 2094  
 QY 2101 CTGACGAGCCATTCAGCTGGCCCTGACGAGACCGGCGAGCGAGGTGAACATCGTGACCGAC 2160  
 Db 2101 CTGACGAGCCATTCAGCTGGCCCTGACGAGACCGGCGAGCGAGGTGAACATCGTGACCGAC 2160  
 QY 2095 CTGACGAGCCATTCAGCTGGCCCTGACGAGACCGGCGAGCGAGGTGAACATCGTGACCGAC 2154  
 Db 2095 CTGACGAGCCATTCAGCTGGCCCTGACGAGACCGGCGAGCGAGGTGAACATCGTGACCGAC 2154  
 QY 2161 AGCCAGTACGCGCTGGGATCATTCAGGCCCGAGCCCGAGAGCGAGCGAGTGGTG 2220  
 Db 2161 AGCCAGTACGCGCTGGGATCATTCAGGCCCGAGCCCGAGAGCGAGCGAGTGGTG 2220  
 QY 2155 AGCCAGTACGCGCTGGGATCATTCAGGCCCGAGCCCGAGAGCGAGCGAGTGGTG 2214  
 Db 2155 AGCCAGTACGCGCTGGGATCATTCAGGCCCGAGCCCGAGAGCGAGCGAGTGGTG 2214  
 QY 2221 AACACAGATCATCGAGAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGTGGCCGCC 2280  
 Db 2221 AACACAGATCATCGAGAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGTGGCCGCC 2280  
 QY 2215 AACACAGATCATCGAGAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGTGGCCGCC 2274  
 Db 2215 AACACAGATCATCGAGAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGTGGCCGCC 2274  
 QY 2281 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGGTGGAGCAAGGGCATCGGCAAG 2340  
 Db 2281 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGGTGGAGCAAGGGCATCGGCAAG 2340  
 QY 2275 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGGTGGAGCAAGGGCATCGGCAAG 2334  
 Db 2275 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGGTGGAGCAAGGGCATCGGCAAG 2334  
 QY 2341 GTGCTGTCTGAGCGGATCGATGGCGGATCGTGATCTTACAGATGATCGAGCAAGCTG 2400  
 Db 2341 GTGCTGTCTGAGCGGATCGATGGCGGATCGTGATCTTACAGATGATCGAGCAAGCTG 2400

Db 2335 GTGCTGTTCTGGACGGCATGATGGCGGCATCGTGATCTACAGTATACGAGCAACCTG 2394  
 Qy 2401 TACGTGGCCAGCGCGCCCTAGGATGATTAAGCTTCCCGGGGTAGCACCGGTGAA 2460  
 Db 2395 TACGTGGCCAGCGCGCCCTAGGATGATTAAGCTTCCCGGGGTAGCACCGGTGAA 2454  
 Qy 2461 TTC 2463  
 Db 2455 TTC 2457  
 RESULT 9  
 ACA03548  
 ID ACA03548 standard; DNA; 2457 BP.  
 AC ACA03548;  
 XX  
 DT 22-MAY-2003 (first entry)  
 XX  
 DE Synthetic DNA encoding immunogenic HIV peptide #31.  
 XX  
 KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 OS Synthetic.  
 XX  
 PN W02003004657-A1.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021421.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX  
 PI Zur Megede J, Barnett SW, Lian Y;  
 XX  
 DR WPI; 2003-221602/21.  
 XX  
 PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX  
 PS Example 1; Fig 36; 262pp; English.  
 XX  
 CC The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide  
 XX  
 SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;  
 Query Match 98.78; Score 2430.2; DB 8; Length 2457;  
 Best Local Similarity 99.68; Pred. No. 4.6e-293;  
 Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
 7 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACCAAGCCCAACATCTCTGATGCGAGCGCAGC 66  
 |||||

Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGGCCACCAAGCCCAACATCTCTGATGCGAGCGCAGC 60  
 Qy 67 AACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGAGGGGCCACATC 126  
 Db 61 AACTTCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGAGGGGCCACATC 120  
 Qy 127 GCCCGCAACTGCGCGCCCCCGCAAGAGGGGTGCTTGAAGTTCGCGCAAGAGGGGCCAC 186  
 Db 121 GCCCGCAACTGCGCGCCCCCGCAAGAGGGGTGCTTGAAGTTCGCGCAAGAGGGGCCAC 180  
 Qy 187 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCC 246  
 Db 181 CAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTTGGCTTCCCC 240  
 Qy 247 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAACCGCGCCCAACAGCCCCACAGCCGC 306  
 Db 241 CAGGGCAAGGCCCGCGAGTTCCCGAGGAGCAACCGCGCCCAACAGCCCCACAGCCGC 300  
 Qy 307 GAGCTGAGGTGCGCGCGCAACACCCCGCAGGAGGCCCGCGCCCGAGCGCCAGGGCACC 366  
 Db 301 GAGCTGAGGTGCGCGCGCAACACCCCGCAGGAGGCCCGCGCCCGAGCGCCAGGGCACC 360  
 Qy 367 CTGAATCTTCCCGCAGATCAACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGC 426  
 Db 361 CTGAATCTTCCCGCAGATCAACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGC 420  
 Qy 427 CAGATCAAGGAGGCCCTGTGTGGACACCGCGCGCGAGCACCTGTGTGAGGAGATGAGC 486  
 Db 421 CAGATCAAGGAGGCCCTGTGTGGACACCGCGCGCGAGCACCTGTGTGAGGAGATGAGC 480  
 Qy 487 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGGCTTTCATCAAGGTGCGC 546  
 Db 481 CTGCCCGGCAAGTGGAAAGCCCAAGATGATCGCGGGCATCGCGGGCTTTCATCAAGGTGCGC 540  
 Qy 547 CAGTACGACAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCTGTGTGATC 606  
 Db 541 CAGTACGACAGATCTGTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCTGTGTGATC 600  
 Qy 607 GGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666  
 Db 601 GGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660  
 Qy 667 AACTTCTCCCATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGACGGC 726  
 Db 661 AACTTCTCCCATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGCATGGACGGC 720  
 Qy 727 CCAGAGTGAAGCAGTGGGCCCTTGAACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 786  
 Db 721 CCAGAGTGAAGCAGTGGGCCCTTGAACCGAGGAGAGATCAAGGCCCTGACCGCATCTGC 780  
 Qy 787 GAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCCCGCGAGAACCCCTTACCAACACC 846  
 Db 781 GAGGAGATGGAGAGGAGGCAAGATCACCAGATCGGCCCGCGAGAACCCCTTACCAACACC 840  
 Qy 847 CCGTGTTCGCCCATCAAGAGAGAGGACAGCACCAGTGGCGCAAGCTGGTGGATTCCTCCG 906  
 Db 841 CCGTGTTCGCCCATCAAGAGAGAGGACAGCACCAGTGGCGCAAGCTGGTGGATTCCTCCG 900  
 Qy 907 GAGCTGAACAGGCGCACCCAGGATCTTGTGGAGGTGAGCTGGGCATCCCCACCCCGCC 966  
 Db 901 GAGCTGAACAGGCGCACCCAGGATCTTGTGGAGGTGAGCTGGGCATCCCCACCCCGCC 960  
 Qy 967 GGCCTGAAGAGAGAGAGAGCGGTGACCGTCTGAGCTGGGCGGACCGCTACTTTCAGGCTG 1026  
 Db 961 GGCCTGAAGAGAGAGAGAGCGGTGACCGTCTGAGCTGGGCGGACCGCTACTTTCAGGCTG 1020  
 Qy 1027 CCCTTGGACGAGGACTTTCGCAAGATACCGCCCTTCAACATCCCCAGCATCAACAACGAG 1086  
 Db 1021 CCCTTGGACGAGGACTTTCGCAAGATACCGCCCTTCAACATCCCCAGCATCAACAACGAG 1080  
 Qy 1087 ACCCCCGGATCGCTTACAGTACAGTGTCTGCCCCCAGGGCTGGAGGGGAGCCCCCAGC 1146  
 Db 1081 ACCCCCGGATCGCTTACAGTACAGTGTCTGCCCCCAGGGCTGGAGGGGAGCCCCCAGC 1140



QY 1147 ATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCCAAACCCCGAGATC 1206  
 Db 1141 ATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCGCCAAACCCCGAGATC 1200  
 QY 1207 GTGATCTACCA-----GSCCCCTCTGTACGTGGGCGAGCGACTGGAGATCGGCCAGCAC 1260  
 Db 1201 GTGATCTACCAAGATCATGACGACCTGTACGTGGGCGAGCGACTGGAGATCGGCCAGCAC 1260  
 QY 1261 CGCGCAAGATCAGAGAGCTGGCAAGCACTGTGCTGCGCTGGGGCTTCAACACCCCGGAC 1320  
 Db 1261 CGCGCAAGATCAGAGAGCTGGCAAGCACTGTGCTGCGCTGGGGCTTCAACACCCCGGAC 1320  
 QY 1321 AAGAGCACCAGAGAGCCCTTCTGTGGATGGGCTACGAGCTGCAACCCCGACAAG 1380  
 Db 1321 AAGAGCACCAGAGAGCCCTTCTGTGGATGGGCTACGAGCTGCAACCCCGACAAG 1380  
 QY 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAG 1440  
 Db 1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAG 1440  
 QY 1441 AAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGATCAAGGTGGCCAG 1500  
 Db 1441 AAGCTGTGGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGGATCAAGGTGGCCAG 1500  
 QY 1501 CTGTGCAAGCTGCTGCGGGGCCCAAGGCCCTGACCGACATGTCGCCCTGACCGAGGAG 1560  
 Db 1501 CTGTGCAAGCTGCTGCGGGGCCCAAGGCCCTGACCGACATGTCGCCCTGACCGAGGAG 1560  
 QY 1561 GCCGAGCTGGAGCTGGCGAGAACCGGAGATCTCTGCGGAGCCCGTGCAGCGCGTGTAC 1620  
 Db 1561 GCCGAGCTGGAGCTGGCGAGAACCGGAGATCTCTGCGGAGCCCGTGCAGCGCGTGTAC 1620  
 QY 1621 TACGACCCCGCAGCAGGACCTGTGTGCGGAGATCCAGAAAGCAGGGCCAGCAGCTGGACC 1680  
 Db 1621 TACGACCCCGCAGCAGGACCTGTGTGCGGAGATCCAGAAAGCAGGGCCAGCAGCTGGACC 1680  
 QY 1681 TACGAGATCTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGC 1740  
 Db 1681 TACGAGATCTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGC 1740  
 QY 1741 ACCGCCACACACAGCTGAGCAGCTGACCGAGCGCGTGCAGAAATGCCCATGAG 1800  
 Db 1741 ACCGCCACACACAGCTGAGCAGCTGACCGAGCGCGTGCAGAAATGCCCATGAG 1800  
 QY 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGGAGACCTGG 1860  
 Db 1801 AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGGAGACCTGG 1860  
 QY 1861 GAGACCTGTGAGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGAGTTGCTGAAC 1920  
 Db 1861 GAGACCTGTGAGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGAGTTGCTGAAC 1920  
 QY 1921 ACCCGCCCTGTTGAGCTGTTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1980  
 Db 1921 ACCCGCCCTGTTGAGCTGTTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAG 1980  
 QY 1981 ACCTTCTACGTGGAGCGCGCCCAACCCGAGACCAAGATCGGCAAGCCGGCTTACGTG 2040  
 Db 1981 ACCTTCTACGTGGAGCGCGCCCAACCCGAGACCAAGATCGGCAAGCCGGCTTACGTG 2040  
 QY 2041 ACCGACCGGGCCCGGCAAGATCGTGTGCTGACCCGAGACCAACCAAGAGACCGAG 2100  
 Db 2041 ACCGACCGGGCCCGGCAAGATCGTGTGCTGACCCGAGACCAACCAAGAGACCGAG 2100  
 QY 2101 CTGACGAGCATTCAGCTGGCCCTGACGACAGCGGACGAGGTGAACATCGTACCGAC 2160  
 Db 2101 CTGACGAGCATTCAGCTGGCCCTGACGACAGCGGACGAGGTGAACATCGTACCGAC 2160  
 QY 2161 AGCCAGTACGCTGGGGCATCATCCAGGCCCGAGCCCGAAGAGCGAGGAGCTGTG 2220  
 Db 2161 AGCCAGTACGCTGGGGCATCATCCAGGCCCGAGCCCGAAGAGCGAGGAGCTGTG 2220

QY 2221 AACGAGATCATCGACAGCTGATCAAGAGGAGAGAGGTGTACTCTGAGCTGGGTGCCGCC 2280  
 Db 2221 AACGAGATCATCGACAGCTGATCAAGAGGAGAGAGGTGTACTCTGAGCTGGGTGCCGCC 2280  
 QY 2281 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGGCAAGGGCATCCGCAAG 2340  
 Db 2281 CACAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGGCAAGGGCATCCGCAAG 2340  
 QY 2341 GTGCTGTTCTTGGAGCGGCATCGATGCGCGCATCGTGTATCTACAGTACATGACGACCTG 2400  
 Db 2341 GTGCTGTTCTTGGAGCGGCATCGATGCGCGCATCGTGTATCTACAGTACATGACGACCTG 2400  
 QY 2401 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457  
 Db 2401 TACGTGGGCGAGCGCGCCCTTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 10  
 ADC13266  
 ID ADC13266 standard; DNA; 2457 BP.  
 AC ADC13266;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE DNA of HIV construct p2Pol-opt\_C SEQ ID NO 45.  
 XX  
 KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 FN WO2003004620-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021420.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (UYST-) UNIV STELLENBOSCH.  
 XX  
 Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 WPI; 2003-221593/21.  
 XX  
 New expression cassette comprising a polynucleotide sequence encoding a  
 polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 Prot, or Rev polypeptide, useful for immunization, or generating  
 packaging cell lines.  
 XX  
 Disclosure; Fig 42; 301pp; English.  
 PS  
 CC The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC subcutaneously, intradermally, intramuscularly, intranasally,  
 CC intrarectally, orally or intravenously. The expression cassette is useful  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C-related sequence of the invention.  
 XX  
 SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 98.7%; Score 2430.2; DB 10; Length 2457;  
 Best Local Similarity 99.6%; Pred. No. 4.6e-293;  
 Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 7 GCACCATGCGCGAGGCGCATGAGCCAGCGCCACAGCGCCCAACATCTCTGATCGAGCGCAGC 66  
DB 1 GCACCATGCGCGAGGCGCATGAGCCAGCGCCACAGCGCCCAACATCTCTGATCGAGCGCAGC 60  
QY 67 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATC 126  
DB 61 AACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATC 120  
QY 127 GCGCGCAACTGCGCGCGCCCCCGCAAGAGGGTGTCTGGAAGTGCAGCAAGGAGGGGCCAC 186  
DB 121 GCGCGCAACTGCGCGCGCCCCCGCAAGAGGGTGTCTGGAAGTGCAGCAAGGAGGGGCCAC 180  
QY 187 CAGATGAAGGACTGCACCCGAGCGCGCAGCCAACTTCTTCCGCGAGGACTGTGGCTTCCCC 246  
DB 181 CAGATGAAGGACTGCACCCGAGCGCGCAGCCAACTTCTTCCGCGAGGACTGTGGCTTCCCC 240  
QY 247 CAGGGCAAGGCCCGCGAGTCTCCCGAGCGAGAGCAAGAACCGCGCCCAACAGCCCCCAAGCGCGC 306  
DB 241 CAGGGCAAGGCCCGCGAGTCTCCCGAGCGAGAGCAAGAACCGCGCCCAACAGCCCCCAAGCGCGC 300  
QY 307 GAGCTGAGGTGCGCGCGGCAACCCCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 366  
DB 301 GAGCTGAGGTGCGCGCGGCAACCCCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
QY 367 CTGAACCTTCCCGCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 426  
DB 361 CTGAACCTTCCCGCAGATCACCTGTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGGC 420  
QY 427 CAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGACACCGTGTGTGGAGGAGATGAGC 486  
DB 421 CAGATCAAGGAGGCCCTGTGTGACACCGCGCGCGAGACACCGTGTGTGGAGGAGATGAGC 480  
QY 487 CTGCCCCGAAGTGGAAAGCCAGATGATCGCGGGCATCGCGCGCTTCATCAAGGTGCGC 546  
DB 481 CTGCCCCGAAGTGGAAAGCCAGATGATCGCGGGCATCGCGCGCTTCATCAAGGTGCGC 540  
QY 547 CAGTACCAACAGATCTCTGATCGAGATCTGCGGAAGAGGCCATCGGCACCGTGTCTATC 606  
DB 541 CAGTACCAACAGATCTCTGATCGAGATCTGCGGAAGAGGCCATCGGCACCGTGTCTATC 600  
QY 607 GGCCCCACCCCGTGAAACATCATCGCGCGCAACATCTGACCCAGCTGGGCTGACCCCTG 666  
DB 601 GGCCCCACCCCGTGAAACATCATCGCGCGCAACATCTGACCCAGCTGGGCTGACCCCTG 660  
QY 667 AACTTCCCCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGGACGCGC 726  
DB 661 AACTTCCCCATCGACCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGGACGCGC 720  
QY 727 CCNAGGTGAAGCAGTGGCCCTTGACCGAGAGAGATCAAGGCCCTTGACCGGCATCTGC 786  
DB 721 CCNAGGTGAAGCAGTGGCCCTTGACCGAGAGAGATCAAGGCCCTTGACCGGCATCTGC 780  
QY 787 GAGGAGATGGAGAGGAGGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTCAACACC 846  
DB 781 GAGGAGATGGAGAGGAGGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTTCAACACC 840  
QY 847 CCGTGTTCGCCATCAAGAGAAGGACAGCACCAAGTGGCGCGCAAGCTGGTGGACTTCCGC 906  
DB 841 CCGTGTTCGCCATCAAGAGAAGGACAGCACCAAGTGGCGCGCAAGCTGGTGGACTTCCGC 900  
QY 907 GAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCCAGCCCCGCC 966  
DB 901 GAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCCAGCCCCGCC 960  
QY 967 GGCTTGAAGAGAGAGAGCGTGACCGTGTGGAGTGGCGCGACCGCTTCTTACGCGTG 1026  
DB 961 GGCTTGAAGAGAGAGAGCGTGACCGTGTGGAGTGGCGCGACCGCTTCTTACGCGTG 1020  
QY 1027 CCGCTGACAGGACTTTCGGAAGTACACCGCTTTCACCATCCCCAGCATCAACACGAG 1086  
DB 1021 CCGCTGACAGGACTTTCGGAAGTACACCGCTTTCACCATCCCCAGCATCAACACGAG 1080

QY 1087 ACCCCGCGCATCCGCTACCAAGTACAAGTGTCTCCCGAGGGCTGGAAAGGCGAGCCCCAGC 1146  
DB 1081 ACCCCGCGCATCCGCTACCAAGTACAAGTGTCTCCCGAGGGCTGGAAGGCGAGCCCCAGC 1140  
QY 1147 ATCTTCCAGAGCAGCATGACCAAGATCTCTGAGGCCCTTCCGCGCGCGCAACCCCGAGATC 1206  
DB 1141 ATCTTCCAGAGCAGCATGACCAAGATCTCTGAGGCCCTTCCGCGCGCGCAACCCCGAGATC 1200  
QY 1207 GTGATCTTACCA-----GGCCCCCTGTAGTGTGGGAGCGACCTTGGAGATCGGCCAGCAC 1260  
DB 1201 GTGATCTTACCAAGTACATGGACGACCTGTACTGTGGGAGCGACCTTGGAGATCGGCCAGCAC 1260  
QY 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTCTGCGCTGGGGCTTCAACAACCCCGAC 1320  
DB 1261 CGCGCCAAAGATCGAGGAGCTGCGCAAGCACCTCTGCGCTGGGGCTTCAACAACCCCGAC 1320  
QY 1321 AAGAGCACAGAAAGAGGCCCTTCTCTGTGTGATGGGCTACGAGCTGCAACCCCGACAAG 1380  
DB 1321 AAGAGCACAGAAAGAGGCCCTTCTCTGTGTGATGGGCTACGAGCTGCAACCCCGACAAG 1380  
QY 1381 TGGACCGTGCAGGCCCATCGAGCTGCCGAGAAAGAGAGCTGGACCGGTGAACGACATCCAG 1440  
DB 1381 TGGACCGTGCAGGCCCATCGAGCTGCCGAGAAAGAGAGCTGGACCGGTGAACGACATCCAG 1440  
QY 1441 AAGCTGTGGGCAAGCTGAACTGTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500  
DB 1441 AAGCTGTGGGCAAGCTGAACTGTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG 1500  
QY 1501 CTGTGCAAGTGTCTGCGCGCGCCAAAGGCCCTGACCGACATCTGTGCCCTTGAACGAGGAG 1560  
DB 1501 CTGTGCAAGTGTCTGCGCGCGCCAAAGGCCCTGACCGACATCTGTGCCCTTGAACGAGGAG 1560  
QY 1561 GCCGAGCTGAGCTGCGCGGAGAACCGCGAGATCTCTGCGGAGCCCGTGCACGGCGTGTAC 1620  
DB 1561 GCCGAGCTGAGCTGCGCGGAGAACCGCGAGATCTCTGCGGAGCCCGTGCACGGCGTGTAC 1620  
QY 1621 TACGACCCCAAGGAGACCTGTGTGGCGCGAGATCCAGAAAGCAGGGCCACAGCAAGTGAACC 1680  
DB 1621 TACGACCCCAAGGAGACCTGTGTGGCGCGAGATCCAGAAAGCAGGGCCACAGCAAGTGAACC 1680  
QY 1681 TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGGC 1740  
DB 1681 TACGAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGGC 1740  
QY 1741 ACCGCCACACCAAGACGCTGAAGCAGCTGACCGAGGCCGTGAGAAAGATCGCCATCGAG 1800  
DB 1741 ACCGCCACACCAAGACGCTGAAGCAGCTGACCGAGGCCGTGAGAAAGATCGCCATCGAG 1800  
QY 1801 AGCATCGTGTCTGTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAGACCTGG 1860  
DB 1801 AGCATCGTGTCTGTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAGACCTGG 1860  
QY 1861 GAGACCTGTGGAACCGACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTCGTGAAC 1920  
DB 1861 GAGACCTGTGGAACCGACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTCGTGAAC 1920  
QY 1921 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCTGGCGCCGAG 1980  
DB 1921 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCTGGCGCCGAG 1980  
QY 1981 ACCTTCTTACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGCAAGGGCGGCTACGTG 2040  
DB 1981 ACCTTCTTACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGCAAGGGCGGCTACGTG 2040  
QY 2041 ACCGACCGGGGCCGCGAAGATCTGTAGCCTGACCGAGACCAACCAAGGAGACCGAG 2100  
DB 2041 ACCGACCGGGGCCGCGAAGATCTGTAGCCTGACCGAGACCAACCAAGGAGACCGAG 2100  
QY 2101 CTCGAGGCCCATTCAGCTGGGCCCTTGCAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAC 2160  
DB 2101 CTCGAGGCCCATTCAGCTGGGCCCTTGCAGGACAGCGGAGCGAGGTGAACATCTGTGACCGAC 2160  
QY 2161 AGCCAGTACGCCCTTGGGCAATCTCCAGGCCCAAGGCCGAGAGCGGAGCTGGTG 2220

Db 2161 AGCCAGTACGCCCTGGGATCATCAGGCCCGACAGCCCGACAGAGGAGGAGCGAGCTGGTG 2220  
 Qy 2221 AACACAGATCATGACGAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGGTGCCCGCC 2280  
 Db 2221 AACACAGATCATGACGAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGGTGCCCGCC 2280  
 Qy 2281 CACAAGGCGATCGGCGGCAACAGAGAGATCGAAGAGTGGTGAGCAAGGGGATCGCGAAG 2340  
 Db 2281 CACAAGGCGATCGGCGGCAACAGAGAGATCGAAGAGTGGTGAGCAAGGGGATCGCGAAG 2340  
 Qy 2341 GTGCTGTTCTGGAGCGGATCGATGGCGGATCGTGATCTACCACTACATGAGACGACCTG 2400  
 Db 2341 GTGCTGTTCTGGAGCGGATCGATGGCGGATCGTGATCTACCACTACATGAGACGACCTG 2400  
 Qy 2401 TACGTGGGAGCGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457  
 Db 2401 TACGTGGGAGCGGCGGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

RESULT 11

ACA03546  
 ID ACA03546 standard; DNA; 2445 BP.  
 AC ACA03546;  
 XX  
 XX 22-MAY-2003 (first entry)  
 XX  
 XX Synthetic DNA encoding immunogenic HIV peptide #29.  
 XX  
 XX Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;  
 KW gene therapy; packaging cell line; humoral immune response;  
 KW cellular immune response; gene delivery vector; DNA immunisation; ds.  
 XX  
 XX Synthetic.  
 XX  
 XX WO2003004657-A1.  
 XX  
 XX 16-JAN-2003.  
 XX  
 XX 05-JUL-2002; 2002WO-US021421.  
 XX  
 XX 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349728P.  
 PR 16-JAN-2002; 2002US-0349793P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 XX (CHIR ) CHIRON CORP.  
 XX  
 XX Zur Megede J, Barnett SW, Lian Y;  
 XX  
 XX WPI; 2003-221602/21.  
 XX  
 XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C  
 PT polypeptides, useful as immunogenic compositions or vaccines for  
 PT generating humoral or cellular immune responses against HIV in a subject,  
 PT especially humans.  
 XX  
 XX Example 1; Fig 34; 262pp; English.  
 XX  
 XX The invention describes a synthetic polynucleotide encoding 2 or more  
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are  
 CC derived from different HIV subtypes. The polynucleotide is useful for  
 CC immunisation, generation of packaging cell lines, or production of HIV  
 CC polypeptides. The polynucleotide and its encoded proteins are useful as  
 CC immunogenic compositions or vaccines for generating humoral or cellular  
 CC immune responses against HIV in a subject, or for inducing neutralising  
 CC antibodies against HIV. The gene delivery vector comprising the  
 CC polynucleotide is also useful for DNA immunisation of, or for generating  
 CC an immune response (e.g. a humoral or cellular immune response) in, a  
 CC subject such as a mammal, particularly a human. This sequence encodes a  
 CC human immunodeficiency virus immunogenic peptide

XX SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;  
 Query Match 98.4%; Score 2422.6; DB 8; Length 2445;  
 Best Local Similarity 99.6%; Pred. No. 4e-292;  
 Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;  
 Qy 7 GCCACCATGGCCGAGGCGCATGAGCCAGCCACAGCGCACAATCTGATGAGGCGCAGC 66  
 Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGCCACAGCGCACAATCTGATGAGGCGCAGC 60  
 Qy 67 AACTTCAAGGGCGCCNAGGCGCATCATCAAGTGTCTTCACTGGGCAAGAGGGGCCACATC 126  
 Db 61 AACTTCAAGGGCGCCNAGGCGCATCATCAAGTGTCTTCACTGGGCAAGAGGGGCCACATC 120  
 Qy 127 GCCCGCAACTCCCGCGCCCGCCGCAAGAGGGCTGTCTGAAAGTGGCGCAAGAGGGGCCAC 186  
 Db 121 GCCCGCAACTCCCGCGCCCGCCGCAAGAGGGCTGTCTGAAAGTGGCGCAAGAGGGGCCAC 180  
 Qy 187 CAGATGAAGGACTGCAACGAGCGCCAGCCAACTTTCTTCCGCGAGGACCTGGCCTTTCCC 246  
 Db 181 CAGATGAAGGACTGCAACGAGCGCCAGGCCAACTTTCTTCCGCGAGGACCTGGCCTTTCCC 240  
 Qy 247 CAGGCGAAGGCGCCGCGAGTTCCCGAGCGAGCAGAACCGCGCCCAACAGCCCCCAGCCGC 306  
 Db 241 CAGGCGAAGGCGCCGCGAGTTCCCGAGCGAGCAGAACCGCGCCCAACAGCCCCCAGCCGC 300  
 Qy 307 GAGCTGCAAGTGGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 366  
 Db 301 GAGCTGCAAGTGGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 360  
 Qy 367 CTGAATTTCCCGCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGTGTGGCGGC 426  
 Db 361 CTGAATTTCCCGCAGATCACCTGTGTGCGAGCGCCCTCTGTGTGAGCATCAAGTGTGGCGGC 420  
 Qy 427 CAGATCAAGGAGGCGCTGTGACACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486  
 Db 421 CAGATCAAGGAGGCGCTGTGCGCCACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 480  
 Qy 487 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGTGTGGC 546  
 Db 481 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGTGTGGC 540  
 Qy 547 CAGTACACCAAGATCTCTGATGAGATCTGCGCGCAAGAGGCGCATCGGCACCGTGTGATC 606  
 Db 541 CAGTACACCAAGATCTCTGATGAGATCTGCGCGCAAGAGGCGCATCGGCACCGTGTGATC 600  
 Qy 607 GGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 666  
 Db 601 GGGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG 660  
 Qy 667 AACTTTCCCGCATCAGCCCGCATCGAGACCGTCCGCGTGAAGCTGAAGCCCGCATGAGCGGC 726  
 Db 661 AACTTTCCCGCATCAGCCCGCATCGAGACCGTCCGCGTGAAGCTGAAGCCCGCATGAGCGGC 720  
 Qy 727 CCAAGGTGAAGCAGTGGCGCCCTGACCCGAGGAGAGATCAAGGCGCTGACCGCCATCTGC 786  
 Db 721 CCAAGGTGAAGCAGTGGCGCCCTGACCCGAGGAGAGATCAAGGCGCTGACCGCCATCTGC 780  
 Qy 787 GAGGAGATGAGAGAGGAGGCGAAGATCACCAAGATCGGCCCGCGAGAACCCCTACAGACC 846  
 Db 781 GAGGAGATGAGAGAGGAGGCGAAGATCACCAAGATCGGCCCGCGAGAACCCCTACAGACC 840  
 Qy 847 CCGGTGTTCCCGCATCAAGAGAGAGGAGCAGCAGCAAGTGGCGCAAGCTGGTGCATTTCCGC 906  
 Db 841 CCGGTGTTCCCGCATCAAGAGAGAGGAGCAGCAGCAAGTGGCGCAAGCTGGTGCATTTCCGC 900  
 Qy 907 GAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGCATCCCCCAGCCCGCC 966  
 Db 901 GAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGCATCCCCCAGCCCGCC 960  
 Qy 967 GGCCTGAAGAGAGAGAGGCGTACCGTGTGTGAGCGTGGGCGAGCGCTTACTTTCAGCGGT 1026

Db 961 GGCCTGAAGAAGAGAGCGTGACCGGTGCTGGACGCTGGCGGACGCGCTACTTTCAGCGTG 1020  
Qy 1027 CCCCTGACGAGGACTTTCGCGAGTACACCGCTTTCACCATCCCGACATCAACACGAG 1086  
Db 1021 CCCCTGACGAGGACTTTCGCGAGTACACCGCTTTCACCATCCCGACATCAACACGAG 1080  
Qy 1087 ACCCCGCGGACTCCGCTACAGTACAAAGTGTGCTGCCCGAGGCTGGAAGGGGAGCGCCGAGC 1146  
Db 1081 ACCCCGCGGACTCCGCTACAGTACAAAGTGTGCTGCCCGAGGCTGGAAGGGGAGCGCCGAGC 1140  
Qy 1147 ATCTTCAGAGCAGCATGACCAAGATCCTTGAGGCCCTTCCGCGCCCGCAACCCCGAGATC 1206  
Db 1141 ATCTTCAGAGCAGCATGACCAAGATCCTTGAGGCCCTTCCGCGCCCGCAACCCCGAGATC 1200  
Qy 1207 GTGATCTACAGGCCCCCTGTAGTGGGACAGCGACCTTGAGATCGGCGACGACCGCGCC 1266  
Db 1201 GTGATCTACAGGCCCCCTGTAGTGGGACAGCGACCTTGAGATCGGCGACGACCGCGCC 1260  
Qy 1267 AAGATCGAGGAGCTGCGCAAGCAGCTGTGGCTGGGGCTTACACACCCCGCAACGAAG 1326  
Db 1261 AAGATCGAGGAGCTGCGCAAGCAGCTGTGGCTGGGGCTTACACACCCCGCAACGAAG 1320  
Qy 1327 CACCAGAGGAGCCCCCTTCTGTGATGGGCTACAGAGTGCACCCCGCAACAGTGGAAC 1386  
Db 1321 CACCAGAGGAGCCCCCTTCTGTGATGGGCTTCTGCCCCAT-----CGAGCTGCACCCCGCAACAGTGGAAC 1374  
Qy 1387 GTGAGGCCATCGAGCTGCCGCGAGAGGAGCTGGACCGTGAACGATCCAGAGCTG 1446  
Db 1375 GTGAGGCCATCGAGCTGCCGCGAGAGGAGCTGGACCGTGAACGATCCAGAGCTG 1434  
Qy 1447 GTGGGCAAGCTGAATCGGGCCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGC 1506  
Db 1435 GTGGGCAAGCTGAATCGGGCCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGC 1494  
Qy 1507 AAGCTGTGCGCGCGCCAAAGGCCCTGACCGACATCGTGCCCTTGACCGAGGAGCGGAG 1566  
Db 1495 AAGCTGTGCGCGCGCCAAAGGCCCTGACCGACATCGTGCCCTTGACCGAGGAGCGGAG 1554  
Qy 1567 CTGAGCTGGCGGAGAACCGGAGATCTTGGCGAGCGCGTGCACGCGGTGTACTAGAC 1626  
Db 1555 CTGAGCTGGCGGAGAACCGGAGATCTTGGCGAGCGCGTGCACGCGGTGTACTAGAC 1614  
Qy 1627 CCAGAGGAGCACTGTGTGGCGGAGATCCAGAGCAGGCGCCACACCAAGTGACCTACAG 1686  
Db 1615 CCAGAGGAGCACTGTGTGGCGGAGATCCAGAGCAGGCGCCACACCAAGTGACCTACAG 1674  
Qy 1687 ATCTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1746  
Db 1675 ATCTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1734  
Qy 1747 CACACCAACGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGGAGCATC 1806  
Db 1735 CACACCAACGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGGAGCATC 1794  
Qy 1807 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGAGCTGGGAGACC 1866  
Db 1795 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGAGCTGGGAGACC 1854  
Qy 1867 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTGTTGTAACACCCCC 1926  
Db 1855 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAGTGGGAGTTGTTGTAACACCCCC 1914  
Qy 1927 CCCCTGTGAGCTGTGTACCACTGAGGAGGAGCCCATCATCGGCGCGGAGACCTTC 1986  
Db 1915 CCCCTGTGAGCTGTGTACCACTGAGGAGGAGCCCATCATCGGCGCGGAGACCTTC 1974  
Qy 1987 TACGTGAGCGGCGCGCCCAACCGGAGACCAAGATCGCGAAGGCGGCTACGTACCCGAC 2046  
Db 1975 TACGTGAGCGGCGCGCCCAACCGGAGACCAAGATCGCGAAGGCGGCTACGTACCCGAC 2034  
Qy 2047 CGGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTGAG 2106  
Db 2035 CGGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCGAGCTGAG 2094

Qy 2107 GCATCCAGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTGACCGACAGCCAG 2166  
Db 2095 GCATCCAGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTGACCGACAGCCAG 2154  
Qy 2167 TACGCCCTGGGCATCATCCAGGCCAGCCCGACAAGAGCGAGAGCTGGTGAACAG 2226  
Db 2155 TACGCCCTGGGCATCATCCAGGCCAGCCCGACAAGAGCGAGAGCTGGTGAACAG 2214  
Qy 2227 ATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGAGCTGGGTGCCGCCACAAG 2286  
Db 2215 ATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGAGCTGGGTGCCGCCACAAG 2274  
Qy 2287 GGATCGCGCGGCAACGAGCAGATCGCAAGCTGTGTAGCAAGGCGCATCCGCAAGGTGCTG 2346  
Db 2275 GGATCGCGCGGCAACGAGCAGATCGCAAGCTGTGTAGCAAGGCGCATCCGCAAGGTGCTG 2334  
Qy 2347 TTCTTGGACGGCATCGATGGCGGCATCGTGTATCTACAGTACATGGACGACCTGTACGTG 2406  
Db 2335 TTCTTGGACGGCATCGATGGCGGCATCGTGTATCTACAGTACATGGACGACCTGTACGTG 2394  
Qy 2407 GGACGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
Db 2395 GGACGCGCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2445

## RESULT 12

ADCL3264

ID ADCL3264 standard; DNA; 2445 BP.

XX AC ADCL3264;

XX AC ADCL3264;

DT 18-DEC-2003 (first entry)

XX DNA of HIV construct p2Pol-opt-YMMW\_C SEQ ID NO 43.

DE expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;

XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX Human immunodeficiency virus.

XX WO200304620-A2.

XX 16-JAN-2003.

XX 05-JUL-2002; 2002WO-US021420.

XX 05-JUL-2001; 2001US-0303192P.

XX 31-AUG-2001; 2001US-0316860P.

XX 16-JAN-2002; 2002US-0349871P.

XX (CHIR ) CHIRON CORP.

XX (UYST-) UNIV STELLENBOSCH.

XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;

XX WPI; 2003-221593/21.

XX New expression cassette comprising a polynucleotide sequence encoding a

XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,

XX Prot, or Rev polypeptide, useful for immunization, or generating

XX packaging cell lines.

XX Disclosure; Fig 40; 301pp; English.

XX The invention relates to a novel expression cassette comprising a

XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,

XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel

XX expression cassette can be used to treat HIV type C by gene therapy or

XX used in the development of a vaccine. The gene delivery vector is

XX administered intramuscularly, intramucosally, intranasally,

XX subcutaneously, intradermally, transdermally, intravaginally,

XX intrarectally, orally or intravenously. The expression cassette is useful

CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.

XX Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 98.4%; Score 2422.6; DB 10; Length 2445;  
 Best Local Similarity 99.6%; Pred. No. 4e-292;  
 Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY	7	GCACCATGCGCGAGCCATGAGCCAGCCAGCGCCCAACATCCTGTATGAGCGCAGC	66
DB	1	GCACCATGCGCGAGCCATGAGCCAGCGCCCAACATCCTGTATGAGCGCAGC	60
QY	67	AACATTCAGAGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGCGCACATC	126
DB	61	AACATTCAGAGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGCGCACATC	120
QY	127	GCCCGCAACTGCGCGCGCCCCCGGCAAGAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCAC	186
DB	121	GCCCGCAACTGCGCGCGCCCCCGGCAAGAGAGGCTGCTGGAAGTGGCGCAAGAGGCGCAC	180
QY	187	CAGATGAAGGACTGCAACGAGCGCGCAGCGCAACTTCTTCGCGAGGACCTGSCCTTCCCC	246
DB	181	CAGATGAAGGACTGCAACGAGCGCGCAGCGCAACTTCTTCGCGAGGACCTGSCCTTCCCC	240
QY	247	CAGGCAAGGCGCGCGAGTTCCCAAGCGAGCAGAAACCGCGCAACAGCCCCCAACAGCGCG	306
DB	241	CAGGCAAGGCGCGCGAGTTCCCAAGCGAGCAGAAACCGCGCAACAGCCCCCAACAGCGCG	300
QY	307	GAGTGTGAGGTGCGCGCGCAACCCCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG	366
DB	301	GAGTGTGAGGTGCGCGCGCAACCCCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG	360
QY	367	CTGAACTTCCCGCAGATCACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	426
DB	361	CTGAACTTCCCGCAGATCACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	420
QY	427	CAGATCAAGGAGGCGCTGCGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	486
DB	421	CAGATCAAGGAGGCGCTGCGACACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	480
QY	487	CTGCCCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAGTGTGCG	546
DB	481	CTGCCCGCAAGTGAAGCCCAAGATGATGCGCGCGCATCGCGCGCTTCAATCAAGTGTGCG	540
QY	547	CAGTACGACGATCTCTGATCGAGATCTGCGCAGAGAGGCGCATCGGCACCGTGTGATC	606
DB	541	CAGTACGACGATCTCTGATCGAGATCTGCGCAGAGAGGCGCATCGGCACCGTGTGATC	600
QY	607	GCGCCCAACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG	666
DB	601	GCGCCCAACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTG	660
QY	667	AACATTCAGCGCCCATCGAGACCGTGTGCGGAGAGTGAAGTGAAGCGCGCATGGAAGCG	726
DB	661	AACATTCAGCGCCCATCGAGACCGTGTGCGGAGAGTGAAGTGAAGCGCGCATGGAAGCG	720
QY	727	CCCAAGGTGAAGAGTGGCGCGCTGACCGAGAGAGATCAAGGCGCTGACCGCATCTGC	786
DB	721	CCCAAGGTGAAGAGTGGCGCGCTGACCGAGAGAGATCAAGGCGCTGACCGCATCTGC	780
QY	787	GAGGAGTGAAGAGGAGGCAAGATCAACAGATCGCGCGCGCGAGAGCCCTTCAACACCC	846
DB	781	GAGGAGTGAAGAGGAGGCAAGATCAACAGATCGCGCGCGCGAGAGCCCTTCAACACCC	840
QY	847	CCCGTGTTCGCAATCAAGAGAGGAGCAGCAACCAAGTGGCGCAAGTGGTGGACTTCCGC	906
DB	841	CCCGTGTTCGCAATCAAGAGAGGAGCAGCAACCAAGTGGCGCAAGTGGTGGACTTCCGC	900
QY	907	GAGTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCCCAACCCCGCC	966
DB	901	GAGTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCCCAACCCCGCC	960

QY	967	GCCCTGAAGAGAGAGAGAGCGTGAACCTGTGTGGACGTGGCGAGCGCCTTACTTCAAGCGTG	1026
DB	961	GCCCTGAAGAGAGAGAGAGCGTGAACCTGTGTGGACGTGGCGAGCGCCTTACTTCAAGCGTG	1020
QY	1027	CCCTTGAAGAGAGAGAGAGCGTGAACCTGTGTGGACGTGGCGAGCGCCTTACTTCAAGCGTG	1086
DB	1021	CCCTTGAAGAGAGAGAGAGCGTGAACCTGTGTGGACGTGGCGAGCGCCTTACTTCAAGCGTG	1080
QY	1087	ACCCCGCGCTACCGTACAGTACAACTGTGTGGCGAGCGCTGGAAGGCGAGCGCCAGC	1146
DB	1081	ACCCCGCGCTACCGTACAGTACAACTGTGTGGCGAGCGCTGGAAGGCGAGCGCCAGC	1140
QY	1147	ATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCGCTTTCGCGCGCCCAACCCCGAGATC	1206
DB	1141	ATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCGCTTTCGCGCGCCCAACCCCGAGATC	1200
QY	1207	GTGATCTACAGGCGCGCGCTGTGTGTGGCGAGCGCTGGAAGGCGAGCGCCAGCGCC	1266
DB	1201	GTGATCTACAGGCGCGCGCTGTGTGTGGCGAGCGCTGGAAGGCGAGCGCCAGCGCC	1260
QY	1267	AAGATCGAGGAGCTGCGCAAGCAGCTGTGTGGCGAGCGCTGGAAGGCGAGCGCCAGC	1326
DB	1261	AAGATCGAGGAGCTGCGCAAGCAGCTGTGTGGCGAGCGCTGGAAGGCGAGCGCCAGC	1320
QY	1327	CACAGAGAGGCGCGCGCTTCTGTGTGGATGGGCTACGAGCTGCGACCCCGACAAAGTGGAC	1386
DB	1321	CACAGAGAGGCGCGCGCTTCTGTGTGGATGGGCTACGAGCTGCGACCCCGACAAAGTGGAC	1374
QY	1387	GTGAGCGCCATCGAGCTGCGCGAGAGAGAGCTGGAACCGTGAAGAGCATCCAGAGCTG	1446
DB	1375	GTGAGCGCCATCGAGCTGCGCGAGAGAGAGCTGGAACCGTGAAGAGCATCCAGAGCTG	1434
QY	1447	GTGGGCAAGCTGAATGCGGCGCAGCAGATCTACCCCGGCATCAAGTGTGCGCGAGCTGTC	1506
DB	1435	GTGGGCAAGCTGAATGCGGCGCAGCAGATCTACCCCGGCATCAAGTGTGCGCGAGCTGTC	1494
QY	1507	AAGTGTGCGCGCGCGCGCGCTGTGACCAATCTGTGCGCGCTGTGACCGAGAGCGCGAG	1566
DB	1495	AAGTGTGCGCGCGCGCGCGCTGTGACCAATCTGTGCGCGCTGTGACCGAGAGCGCGAG	1554
QY	1567	CTGAGCTGCGCGAGAGCGCGAGATCTGTGCGCGCGCGCTGTGACCGAGAGCTGTACTAC	1626
DB	1555	CTGAGCTGCGCGAGAGCGCGAGATCTGTGCGCGCGCGCTGTGACCGAGAGCTGTACTAC	1614
QY	1627	CCAGAGAGGAGCTGTGTGGCGAGATCCAGAGAGAGGCGCACGACAGTGTGACCTTACAG	1686
DB	1615	CCAGAGAGGAGCTGTGTGGCGAGATCCAGAGAGAGGCGCACGACAGTGTGACCTTACAG	1674
QY	1687	ATCTACAGAGGCGCGCTTCAAGAACCTGGAAGACCGCGCAAGTACGCGCAAGTGGCACCG	1746
DB	1675	ATCTACAGAGGCGCGCTTCAAGAACCTGGAAGACCGCGCAAGTACGCGCAAGTGGCACCG	1734
QY	1747	CACACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGTGAGAGATCGCCATCGAGAGCATC	1806
DB	1735	CACACCAAGAGCTGAAGCAGCTGACCGAGCGCGTGTGAGAGATCGCCATCGAGAGCATC	1794
QY	1807	GTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGGAGACCTGGGAGAC	1866
DB	1795	GTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGGAGACCTGGGAGAC	1854
QY	1867	TGGTGGAGCGACTACTTGGCAGCGCACCTGTGATCTCCCGAGTGGGAGTTCGTGAACACCC	1926
DB	1855	TGGTGGAGCGACTACTTGGCAGCGCACCTGTGATCTCCCGAGTGGGAGTTCGTGAACACCC	1914
QY	1927	CCCTGTGTGAAGCTGTGTGTACAGCTGGAAGAGGCGCCATCATCGGCGCGAGAGCTTTC	1986
DB	1915	CCCTGTGTGAAGCTGTGTGTACAGCTGGAAGAGGCGCCATCATCGGCGCGAGAGCTTTC	1974
QY	1987	TACGTGAAGCGCGCGCGCAACCGCGAGAGCAGAGTGGCAAGGCGCGCTAGTGAACCGAC	2046
DB	1975	TACGTGAAGCGCGCGCGCGCAACCGCGAGAGCAGAGTGGCAAGGCGCGCTAGTGAACCGAC	2034



QY 914 ACAAGCCAGCCAGAGACTTCTGGAGGTGAGCTGGGATCCCCACCCCGCGGCTGA 973  
 Db 2387 ACAGCGCAGCCAGGACTTCTGGAGGTGAGCTGGGATCCCCACCCCGCGGCTGA 2446  
 QY 974 AGAAGAGAGAGCTGACCGTCTGGACGTGGGAGCGCTTACTTTCAGCGTGGCCCTGG 1033  
 Db 2447 AGAAGAGAGAGCTGACCGTCTGGACGTGGGAGCGCTTACTTTCAGCGTGGCCCTGG 2506  
 QY 1034 ACAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACCAACAGAGACCCCG 1093  
 Db 2507 ACAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACCAACAGAGACCCCG 2566  
 QY 1094 GCATCCGCTACAGTACAGCTGTGCCCCAGGGCTGGAAGGCGAGCCCGAGCATCTTCC 1153  
 Db 2567 GCATCCGCTACAGTACAGCTGTGCCCCAGGGCTGGAAGGCGAGCCCGAGCATCTTCC 2626  
 QY 1154 AGAGCAGCATCAACCAAGATCTTGAGCGCTTCCGCGCCCGCAACCCCGAGATCTGATCT 1213  
 Db 2627 AGAGCAGCATCAACCAAGATCTTGAGCGCTTCCGCGCCCGCAACCCCGAGATCTGATCT 2686  
 QY 1214 ACCAGGCCCCCTGTACGTGGGCGAGCACTTGGAGATCGGCAGCACCGCGCCCAAGATCG 1273  
 Db 2687 ACCAGGCCCCCTGTACGTGGGCGAGCACTTGGAGATCGGCAGCACCGCGCCCAAGATCG 2746  
 QY 1274 AGGAGCTGGCGAAGCACTGTGCGCTGGGGTTTCAACCCCGCAAGAGCAACAGA 1333  
 Db 2747 AGGAGCTGGCGAAGCACTGTGCGCTGGGGTTTCAACCCCGCAAGAGCAACAGA 2806  
 QY 1334 AGGAGCCCCCTTCTGTGGATGGGTACGAGCTGCACCCCGCAAGTGGACCGTGCAGC 1393  
 Db 2807 AGGAGCCCCCTTCTGTGGATGGGTACGAGCTGCACCCCGCAAGTGGACCGTGCAGC 2860  
 QY 1394 CCATCGAGCTGCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGGTGGCA 1453  
 Db 2861 CCATCGAGCTGCCGAGAGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGGTGGCA 2920  
 QY 1454 AGCTGAAGCTGGGCGAGAGATCTACCCCGGATCAAGGTGGCGGAGCTGTGCAAGTGC 1513  
 Db 2921 AGCTGAAGCTGGGCGAGAGATCTACCCCGGATCAAGGTGGCGGAGCTGTGCAAGTGC 2980  
 QY 1514 TGCGCGGCGCAAGGCGCTGACGACATCGTCCCTGACCGAGAGGCGCGAGCTGGAGC 1573  
 Db 2981 TGCGCGGCGCAAGGCGCTGACGACATCGTCCCTGACCGAGAGGCGCGAGCTGGAGC 3040  
 QY 1574 TGCGCGGAGAACCGGCGAGATCTTGGCGGAGCGCGTGTACTACGACCCCAAGCA 1633  
 Db 3041 TGCGCGGAGAACCGGCGAGATCTTGGCGGAGCGCGTGTACTACGACCCCAAGCA 3100  
 QY 1634 AGGACCTGGTGGCGAGATCCAGAAAGCGGCGCAGCACTGAGCTTACAGATCTTACC 1693  
 Db 3101 AGGACCTGGTGGCGAGATCCAGAAAGCGGCGCAGCACTGAGCTTACAGATCTTACC 3160  
 QY 1694 AGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCAAGATCGGACCGCCCAACCA 1753  
 Db 3161 AGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCAAGATCGGACCGCCCAACCA 3220  
 QY 1754 ACAGCTGTAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATCGAGAGCATCTGATCT 1813  
 Db 3221 ACAGCTGTAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATCGAGAGCATCTGATCT 3280  
 QY 1814 GGGGCAAGACCCCAAGTTCGCTGCGCCATCCAGAAAGGAGACCTGGAGACCTGGTGA 1873  
 Db 3281 GGGGCAAGACCCCAAGTTCGCTGCGCCATCCAGAAAGGAGACCTGGAGACCTGGTGA 3340  
 QY 1874 CCGACTACTGGCAGGCGCACTGGATCCCGAGTGGAGTTCGTGAACCCCGCCCTGG 1933  
 Db 3341 CCGACTACTGGCAGGCGCACTGGATCCCGAGTGGAGTTCGTGAACCCCGCCCTGG 3400  
 QY 1934 TGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGG 1993  
 Db 3401 TGAAGCTGTGTACAGCTGAGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGG 3460  
 QY 1994 ACGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGTGACCGCGGCGCC 2053

Db 3461 ACGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGCGCC 3520  
 QY 2054 GGCAGAGATCTGAGCTGAGCTGAGCCAGAGACCAACCAAGAGACCGAGCTGCAGGGCATCC 2113  
 Db 3521 GGCAGAGATCTGAGCTGAGCTGAGCCAGAGACCAACCAAGAGACCGAGCTGCAGGGCATCC 3580  
 QY 2114 AGCTGGCCCTCCAGGACAGCGGCGAGCGAGGTGAACATCGTGAACCGAGGATACGCC 2173  
 Db 3581 AGCTGGCCCTCCAGGACAGCGGCGAGCGAGGTGAACATCGTGAACCGAGGATACGCC 3640  
 QY 2174 TGGGCATCATCCAGGCGCGAGCCCGCAAGAGCGGAGAGCTGGTGAACCGAGATACGC 2233  
 Db 3641 TGGGCATCATCCAGGCGCGAGCCCGCAAGAGCGGAGAGCTGGTGAACCGAGATACGC 3700  
 QY 2234 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAAGGGCATCG 2293  
 Db 3701 AGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAAGGGCATCG 3760  
 QY 2294 GCGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATCCGCAAGGTGCTGTTCCTGG 2353  
 Db 3761 GCGCAACGAGCAGATCGACAAAGCTGGTGAAGAGGCGATCCGCAAGGTGCTGTTCCTGG 3820  
 QY 2354 ACGGCATCATGGCGGCGATCTGATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 2413  
 Db 3821 ACGGCATCATGGCGGCGATCTGATCTACAGTACATGAGCAGCTGTACGTGGGCGAGC 3880  
 QY 2414 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
 Db 3881 GCGGCCCTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 14

ADCL13231  
 ID ADCL13231 standard; DNA; 3930 BP.

XX AC ADCL13231;

XX DT 18-DEC-2003 (first entry)

XX DE DNA of HIV construct GagCompIPolmutAtt\_C SEQ ID NO 10.

XX KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.

XX OS Human immunodeficiency virus.

XX FN WO2003004620-A2.

XX PD 16-JAN-2003.

XX PF 05-JUL-2002; 2002WO-US021420.

XX PR 05-JUL-2001; 2001US-0303192P.

XX PR 31-AUG-2001; 2001US-0316860P.

XX PR 16-JAN-2002; 2002US-0349871P.

XX PA (CHIR ) CHIRON CORP.

XX PA (UYST-) UNIV STELLENBOSCH.

XX PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ;

XX DR WPI; 2003-221593/21.

XX PT New expression cassette comprising a polynucleotide sequence encoding a  
 polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 Prot, or Rev polypeptide, useful for immunization, or generating  
 packaging cell lines.

XX PS Disclosure; Fig 7; 301pp; English.

XX CC The invention relates to a novel expression cassette comprising a  
 polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,





Db 3401 TGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGCGCCGAGACCTTCTACGTGG 3460  
 Qy 1994 ACGGCGCCGCAACCGCAGAGCAAGATCGGCAAGCCCGGTACTCGTACCGACCGGGGCC 2053  
 Db 3461 ACGGCGCCGCAACCGCAGAGCAAGATCGGCAAGCCCGGTACTCGTACCGACCGGGGCC 3520  
 Qy 2054 GGCAGAAATCGTGTAGCTGTACCGAGACCAACCAAGACCGAGCTGCAGGCGCATCC 2113  
 Db 3521 GGCAGAAATCGTGTAGCTGTACCGAGACCAACCAAGACCGAGCTGCAGGCGCATCC 3580  
 Qy 2114 AGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTACCGACAGCCAGTACGCCC 2173  
 Db 3581 AGCTGGCCCTGCAGGACAGCGGACGAGGTGAACATCGTACCGACAGCCAGTACGCCC 3640  
 Qy 2174 TGGGCATCATCCAGGCCCGCCGCAAGAGCGAGCGAGCTGGTGAACAGATCATCG 2233  
 Db 3641 TGGGCATCATCCAGGCCCGCCGCAAGAGCGAGCGAGCTGGTGAACAGATCATCG 3700  
 Qy 2234 AGCAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGGGTGGCCGCCACAAAGGCGCATCG 2293  
 Db 3701 AGCAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGGGTGGCCGCCACAAAGGCGCATCG 3760  
 Qy 2294 GCGGCAAGCAGCAGATCGACAAGCTGGTGAGCAAGGCGATCGCAAGGTGCTGTTCCTGG 2353  
 Db 3761 GCGGCAAGCAGCAGATCGACAAGCTGGTGAGCAAGGCGATCGCAAGGTGCTGTTCCTGG 3820  
 Qy 2354 ACGGCATCGATGGCGGCGCATCGTGATCTACCAAGTACATGGACGACCTGTACGCGGCGCG 2413  
 Db 3821 ACGGCATCGATGGCGGCGCATCGTGATCTACCAAGTACATGGACGACCTGTACGCGGCGCG 3880  
 Qy 2414 GCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 2457  
 Db 3881 GCGGCCCTTAGGATCGATTAAAGCTTCCCGGGGCTAGCACCGGT 3924

RESULT 15  
 ADCL3232  
 ID ADCL3232 standard; DNA; 3930 BP.  
 XX  
 AC ADCL3232;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE DNA of HIV construct GagComplPolmutIna\_C SEQ ID NO 11.  
 XX  
 KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
 KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 PN WO2003004620-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 05-JUL-2002; 2002WO-US021420.  
 XX  
 PR 05-JUL-2001; 2001US-0303192P.  
 PR 31-AUG-2001; 2001US-0316860P.  
 PR 16-JAN-2002; 2002US-0349871P.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (UYST-) UNIV STELLENBOSCH.  
 XX  
 Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;  
 WPI; 2003-221593/21.  
 XX  
 New expression cassette comprising a polynucleotide sequence encoding a  
 polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
 Prot, or Rev polypeptide, useful for immunization, or generating  
 packaging cell lines.  
 XX  
 Disclosure; Fig 8; 301pp; English.

XX The invention relates to a novel expression cassette comprising a  
 CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,  
 CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel  
 CC expression cassette can be used to treat HIV type C by gene therapy or  
 CC used in the development of a vaccine. The gene delivery vector is  
 CC administered intramuscularly, intravenously, intranasally,  
 CC subcutaneously, intradermally, transdermally, or intravenously. The  
 CC for immunisation, generating packaging cell lines and producing HIV  
 CC polypeptides. This polynucleotide sequence represents the DNA of an HIV  
 CC Type C related sequence of the invention.  
 XX  
 SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;  
 Query Match 98.0%; Score 2414; DB 10; Length 3930;  
 Best Local Similarity 99.5%; Pred. No. 4.e-291;  
 Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;  
 Qy 14 TGGCGAGGCGCATGAGCCAGGCCACCGCGCAACATCTGTATGAGCGCGCAACTTCA 73  
 Db 1487 TCGCGAGGCGCATGAGCCAGGCCACCGCGCAACATCTGTATGAGCGCGCAACTTCA 1546  
 Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGCGGCACATCGCCGCA 133  
 Db 1547 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTCGGCAAGGAGGCGGCACATCGCCGCA 1606  
 Qy 134 ACTGCGCGCGCCCCCGCAAGAGGCGTCTCGAAGTGCAGAGGAGGCGGCACACAGATGA 193  
 Db 1607 ACTGCGCGCGCCCCCGCAAGAGGCGTCTCGAAGTGCAGAGGAGGCGGCACACAGATGA 1666  
 Qy 194 AGGACTGACCGAGCGCGCGCGCAACTTCTTCGCGAGGACCTGGCTTCCCGCCAGGCA 253  
 Db 1667 AGGACTGACCGAGCGCGCGCGCAACTTCTTCGCGAGGACCTGGCTTCCCGCCAGGCA 1726  
 Qy 254 AGGGCGCGAGTTCGCCAGCGAGCAGACACCGCGCCACAGCGCCACACAGCGCGGAGTGC 313  
 Db 1727 AGGGCGCGAGTTCGCCAGCGAGCAGACACCGCGCCACAGCGCCACACAGCGCGGAGTGC 1786  
 Qy 314 AGGTGCGCGCGCAACACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373  
 Db 1787 AGGTGCGCGCGCAACACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846  
 Qy 374 TCCCCCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433  
 Db 1847 TCCCCCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906  
 Qy 434 AGGAGGCGCTGCTGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493  
 Db 1907 AGGAGGCGCTGCTGGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966  
 Qy 494 GCAAGTGGAGGCCAAGATGATCGCGCGCATCGCGCGCTTCAAGGTGGCGCGCGCGCGCG 553  
 Db 1967 GCAAGTGGAGGCCAAGATGATGATCGCGCGCATCGCGCGCTTCAAGGTGGCGCGCGCG 2026  
 Qy 554 ACCAGATCTGTATCGAGATCTGCGCGCAAGAGGCCATCGGCAACCGTGTGTATCGCGCG 613  
 Db 2027 ACCAGATCTGTATCGAGATCTGCGCGCAAGAGGCCATCGGCAACCGTGTGTATCGCGCG 2086  
 Qy 614 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGACTTCC 673  
 Db 2087 CCCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGACTTCC 2146  
 Qy 674 CCATCAGGCCCATCGAGACCGTGGCGGTGAAGCCCGGCGATGAGCGCGCGCGCGCGCG 733  
 Db 2147 CCATCAGGCCCATCGAGACCGTGGCGGTGAAGCCCGGCGATGAGCGCGCGCGCGCGCG 2206  
 Qy 734 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 793  
 Db 2207 TGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGA 2266  
 Qy 794 TGGAGAGGAGGCGAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853

Db 2267 TGGAGAGGAGGGCAAGATCAACAAGATCGCGCCCGGAGAACCCCTCAACAACACCCCGTGT 2326  
Qy 854 TCGCCATCAAGAGAGGACAGCACCAAGTGGCGCAGCTGTGGATTCGCGAGCTGA 913  
Db 2327 TCGCCATCAAGAGAGGACAGCACCAAGTGGCGCAGCTGTGGATTCGCGAGCTGA 2386  
Qy 914 ACAAGCCACCCAGGACTTCTGGAGAGTGACAGCTGGGCATCCCCACCCCGCGGCGCTGA 973  
Db 2387 ACAAGCCACCCAGGACTTCTGGAGAGTGACAGCTGGGCATCCCCACCCCGCGGCGCTGA 2446  
Qy 974 AGAAGAGAGAGCGTGCACCGTGTGACCTGGGCGCAGCTACTTTCAGCGTGCCTGG 1033  
Db 2447 AGAAGAGAGAGCGTGCACCGTGTGACCTGGGCGCAGCTACTTTCAGCGTGCCTGG 2506  
Qy 1034 ACGAGGACTTTCGCAAGTACACCGCTTTCACCTCCCGCAGCATCAACAAGAGACCCCG 1093  
Db 2507 ACGAGGACTTTCGCAAGTACACCGCTTTCACCTCCCGCAGCATCAACAAGAGACCCCG 2566  
Qy 1094 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCAAGGCTGGAAAGGCGAGCCCGCAGCATCTTCC 1153  
Db 2567 GCATCCGCTACCAAGTACAAAGTGTGCTGCCCAAGGCTGGAAAGGCGAGCCCGCAGCATCTTCC 2626  
Qy 1154 AGAGCAGCATGAACAAGATCTTGAGCCCTTTCGCGCCCGCAACCCCGAGATCGTGATCT 1213  
Db 2627 AGAGCAGCATGAACAAGATCTTGAGCCCTTTCGCGCCCGCAACCCCGAGATCGTGATCT 2686  
Qy 1214 ACCAGGCCCTTGTAGTGGGAGCGACCTTGGAGATCGGCCAGCACCGGCCCAAGATCG 1273  
Db 2687 ACCAGGCCCTTGTAGTGGGAGCGACCTTGGAGATCGGCCAGCACCGGCCCAAGATCG 2746  
Qy 1274 AGAGCTGCGCAAGCAGCTGCTGGCTGGGGCTTTCACCAACCCCGCAAGAAGCACCAGA 1333  
Db 2747 AGAGCTGCGCAAGCAGCTGCTGGCTGGGGCTTTCACCAACCCCGCAAGAAGCACCAGA 2806  
Qy 1334 AGAGCCCTTCTCTGTGATGGGCTTACGAGCTGCAACCCCGCAAGTGGACCGTGACG 1393  
Db 2807 AGAGCCCTTCTCTGTGATGGGCTTTCGAGCTGCAACCCCGCAAGTGGACCGTGACG 2860  
Qy 1394 CCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGGACGACATCCAGAACCTGGTGGGCA 1453  
Db 2861 CCATCGAGCTGCCCGAGAGAGAGCTGGACCGTGGACGACATCCAGAACCTGGTGGGCA 2920  
Qy 1454 AGCTGAATCGGGCCAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 1513  
Db 2921 AGCTGAATCGGGCCAGCCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGC 2980  
Qy 1514 TGGCGCGCGCAAGGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCGCGAGCTGGAGC 1573  
Db 2981 TGGCGCGCGCAAGGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCGCGAGCTGGAGC 3040  
Qy 1574 TGGCCGAGAACCGCGAGATCTTGGCGAGCCCGTGGACGCGGTGTACTAGACCCGAGCA 1633  
Db 3041 TGGCCGAGAACCGCGAGATCTTGGCGAGCCCGTGGACGCGGTGTACTAGACCCGAGCA 3100  
Qy 1634 AGGACCTGGTGGCGAGATCCAGAAAGCAGGCGCCACGACAGTGGACCTTACGATCTACC 1693  
Db 3101 AGGACCTGGTGGCGAGATCCAGAAAGCAGGCGCCACGACAGTGGACCTTACGATCTACC 3160  
Qy 1694 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCA 1753  
Db 3161 AGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCA 3220  
Qy 1754 ACGACGTGAAGCAGCTGACCGAGCCCGTGCAGAGATCGGCATGGAGAGCATCGTGATCT 1813  
Db 3221 ACGACGTGAAGCAGCTGACCGAGCCCGTGCAGAGATCGGCATGGAGAGCATCGTGATCT 3280  
Qy 1814 GGGCAAGACCCCAAGTTCGCTCGCTCCATCCAGAAAGGAGACCTGGGAGACCTTGGTGA 1873  
Db 3281 GGGCAAGACCCCAAGTTCGCTCGCTCCATCCAGAAAGGAGACCTGGGAGACCTTGGTGA 3340  
Qy 1874 CCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG 1933  
Db 3341 CCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGG 3400

Qy 1934 TGAAGCTGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACCTGG 1993  
Db 3401 TGAAGCTGTGTTACAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACCTGG 3460  
Qy 1994 ACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGAACGACCGGGGCC 2053  
Db 3461 ACGGCGCGCCAAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGAACGACCGGGGCC 3520  
Qy 2054 GGCAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAAAGACGAGCTGCAGGCCATCC 2113  
Db 3521 GGCAGAAAGATCGTGAGCTGACCGAGACCAACCAAGAAAGACGAGCTGCAGGCCATCC 3580  
Qy 2114 AGCTGGCCCTGCAAGGACAGCGGCGAGGCTGAACATCGTGACCGACAGCCAGTACGCC 2173  
Db 3581 AGCTGGCCCTGCAAGGACAGCGGCGAGGCTGAACATCGTGACCGACAGCCAGTACGCC 3640  
Qy 2174 TGGGCATCATCCAGGCCCAAGCCGACAAAGAGCGAGCTGGTGAACCAAGGGCATCG 2233  
Db 3641 TGGGCATCATCCAGGCCCAAGCCGACAAAGAGCGAGCTGGTGAACCAAGGGCATCG 3700  
Qy 2234 AGCAGCTGATCAAGAAAGGAGAGGTTGATCCTTGAGCTGGGTGCCGCCCAAGGGCATCG 2293  
Db 3701 AGCAGCTGATCAAGAAAGGAGAGGTTGATCCTTGAGCTGGGTGCCGCCCAAGGGCATCG 3760  
Qy 2294 GCGGCAACGAGCAGATCGCAAGCTGGTGAGCAAGGCGATCCGCAAGGTGCTTCTCTGG 2353  
Db 3761 GCGGCAACGAGCAGATCGCAAGCTGGTGAGCAAGGCGATCCGCAAGGTGCTTCTCTGG 3820  
Qy 2354 ACGGCATCGATGGCGGCATCGTGATCTTACAGTACATGGACGACCTGTACTGGGCAAGCG 2413  
Db 3821 ACGGCATCGATGGCGGCATCGTGATCTTACAGTACATGGACGACCTGTACTGGGCAAGCG 3880  
Qy 2414 GCGGCCCTTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 2457  
Db 3881 GCGGCCCTTAGGATCGATTAAAAGCTTCCCGGGGCTAGCACCGGT 3924

Search completed: June 1, 2005, 11:33:26  
Job time : 1275.33 secs



Db 181 GGCACACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGAACTTGCC 240  
Qy 241 TTCCCCAGAGGAGCCCGAGTTCCCAAGCAGCAGAACCGCGCCCAAGCCCCCACC 300  
Db 241 TTCCCCAGAGGAGCCCGAGTTCCCAAGCAGCAGAACCGCGCCCAAGCCCCCACC 300  
Qy 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAAACCCCGCAGCGCGCGCGCGCGCGCGCGCG 360  
Db 301 AGCCGCGAGCTGCAGGTGCGCGGCGACAAACCCCGCAGCGCGCGCGCGCGCGCGCGCG 360  
Qy 361 GGCAACCTGAACTTCCCCAGATCAACCTGTGGCAGCGCCCCCTGTGGATCAAGGTG 420  
Db 361 GGCAACCTGAACTTCCCCAGATCAACCTGTGGCAGCGCCCCCTGTGGATCAAGGTG 420  
Qy 421 GGCGGCGAGATCAAGGAGGCCCTGTGACACCGCGCGACACCGTGTGAGGAG 480  
Db 421 GGCGGCGAGATCAAGGAGGCCCTGTGACACCGCGCGACACCGTGTGAGGAG 480  
Qy 481 ATGAGCCTTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGGCTTCATCAAG 540  
Db 481 ATGAGCCTTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGGCTTCATCAAG 540  
Qy 541 GTGCGCGAGTACGACAGATCCTGATCGAGATCTGGCGCAAGAGGCGCATCGGCAACGCTG 600  
Db 541 GTGCGCGAGTACGACAGATCCTGATCGAGATCTGGCGCAAGAGGCGCATCGGCAACGCTG 600  
Qy 601 CTGATCGGCCACCCCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Db 601 CTGATCGGCCACCCCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
Qy 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGAAAGCTGAAAGCCCGGCATG 720  
Db 661 ACCCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGAAAGCTGAAAGCCCGGCATG 720  
Qy 721 GACGGCCCAAGGTGAAGTGGCTGCGCTGACCGAGAGAGATCAAGGCGCTTGAACCGCC 780  
Db 721 GACGGCCCAAGGTGAAGTGGCTGCGCTGACCGAGAGAGATCAAGGCGCTTGAACCGCC 780  
Qy 781 ATCTGCGAGGAGTGAAGAGGAGGCGCAAGATCAACAGATCGCGCCCGAGAACCCCTAC 840  
Db 781 ATCTGCGAGGAGTGAAGAGGAGGCGCAAGATCAACAGATCGCGCCCGAGAACCCCTAC 840  
Qy 841 AACACCCCGGTGTTCCGCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Db 841 AACACCCCGGTGTTCCGCATCAAGAAAGAGACAGCACCAAGTGGCGCAAGCTGGTGGAC 900  
Qy 901 TTCCGCGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGCGATCCCCAC 960  
Db 901 TTCCGCGAGCTGAACAAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGCGATCCCCAC 960  
Qy 961 CCGCGCGCTGAAGAGAGAGAGCGTGACCGTGTGGAGCGCGCTACTTTC 1020  
Db 961 CCGCGCGCTGAAGAGAGAGAGCGTGACCGTGTGGAGCGCGCTACTTTC 1020  
Qy 1021 AGCGTGCCTTGGACAGGACTTCCGCAAGTACACCGCTTTCACCATCCCAGCATCAAC 1080  
Db 1021 AGCGTGCCTTGGACAGGACTTCCGCAAGTACACCGCTTTCACCATCCCAGCATCAAC 1080  
Qy 1081 AACGAGACCCCGCGCATCCGTTACAGTACAACTGTGTGCCCCAGGGCTGAAGGGCAGC 1140  
Db 1081 AACGAGACCCCGCGCATCCGTTACAGTACAACTGTGTGCCCCAGGGCTGAAGGGCAGC 1140  
Qy 1141 CCGAGCATCTTCAGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200  
Db 1141 CCGAGCATCTTCAGAGCAGCATGACCAAGATCTGGAGCCCTTCGCGCCCGCAACCCC 1200  
Qy 1201 GAGATCGTGTACTACAGGCCCTGTGACGTGGCGAGCGACCTGGAGATCGGCGAGCAC 1260  
Db 1201 GAGATCGTGTACTACAGGCCCTGTGACGTGGCGAGCGACCTGGAGATCGGCGAGCAC 1260  
Qy 1261 CGCGCCCAAGATCGAGGAGCTGCGCAAGCACTGTGTGGCTGGGCTTCAACACCCCGCAC 1320  
Db 1261 CGCGCCCAAGATCGAGGAGCTGCGCAAGCACTGTGTGGCTGGGCTTCAACACCCCGCAC 1320

Qy 1321 AAGAAGCACAGAAGGAGCCCCCTTCTCTGGATGGCTTACGAGCTGCACCCCGACAAG 1380  
Db 1321 AAGAAGCACAGAAGGAGCCCCCTTCTCTGGATGGCTTACGAGCTGCACCCCGACAAG 1380  
Qy 1381 TGGAACCGTGCAGGCCATTCGAGCTGCCGAGAAAGAGAGCTGGACCGTGAACGACATCCAG 1440  
Db 1381 TGGAACCGTGCAGGCCATTCGAGCTGCCGAGAAAGAGAGCTGGACCGTGAACGACATCCAG 1440  
Qy 1441 AAGCTGTGGGCGAGCTGAACCTGGGCCAGCAGATCTACCCCGCATCAAGGTGGCCAG 1500  
Db 1441 AAGCTGTGGGCGAGCTGAACCTGGGCCAGCAGATCTACCCCGCATCAAGGTGGCCAG 1500  
Qy 1501 CTGTGAAGCTGTGTGCGCGCGCAAGGCCCTGACCGACATCTGTGCCCTGACCGAGGAG 1560  
Db 1501 CTGTGAAGCTGTGTGCGCGCGCAAGGCCCTGACCGACATCTGTGCCCTGACCGAGGAG 1560  
Qy 1561 GCGAGCTGAGCTGGCCGAGAAACCGCGAGATCTGCGGAGCCGTGCAAGGTGTAC 1620  
Db 1561 GCGAGCTGAGCTGGCCGAGAAACCGCGAGATCTGCGGAGCCGTGCAAGGTGTAC 1620  
Qy 1621 TAGACCCAGCAGAGGACCTGTGGCCGAGATCCAGAAGCAGGGCCACGACCTAGTGACC 1680  
Db 1621 TAGACCCAGCAGAGGACCTGTGGCCGAGATCCAGAAGCAGGGCCACGACCTAGTGACC 1680  
Qy 1681 TACCAGATCTACAGGAGGCCCTTCAAGAACTTGAAGACCGCGCAAGTACGCGCAAGTGGC 1740  
Db 1681 TACCAGATCTACAGGAGGCCCTTCAAGAACTTGAAGACCGCGCAAGTACGCGCAAGTGGC 1740  
Qy 1741 ACCGCCCAACCAACGACGTGAAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAG 1800  
Db 1741 ACCGCCCAACCAACGACGTGAAGCTGACCGAGGCCGTGCAAGAGATCGCCATGGAG 1800  
Qy 1801 AGCATCGTGTCTGGGCGAGACCCCAAGTTCGCGCTGCCATCCAGAGAGGAGACCTGG 1860  
Db 1801 AGCATCGTGTCTGGGCGAGACCCCAAGTTCGCGCTGCCATCCAGAGAGGAGACCTGG 1860  
Qy 1861 GAGACCTGTGTGACCGCACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTAAC 1920  
Db 1861 GAGACCTGTGTGACCGCACTACTGCGAGGCCACCTTGGATCCCGAGTGGGAGTTCTGTAAC 1920  
Qy 1921 ACCCCCCCTGTGTGAAGCTGTGGTACCAAGTGAAGAGGCCCATCATCGGCGCGAG 1980  
Db 1921 ACCCCCCCTGTGTGAAGCTGTGGTACCAAGTGAAGAGGCCCATCATCGGCGCGAG 1980  
Qy 1981 ACCTTCTACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTAGCTG 2040  
Db 1981 ACCTTCTACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTAGCTG 2040  
Qy 2041 ACCGACCGGCGCGCGCAAGAGATCGTGAGCTTGACCGAGACCAACCAAGAGAGCCGAG 2100  
Db 2041 ACCGACCGGCGCGCGCAAGAGATCGTGAGCTTGACCGAGACCAACCAAGAGAGCCGAG 2100  
Qy 2101 CTGAGGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGCGGTGAACATCTGTACCGAC 2160  
Db 2101 CTGAGGCCATTCAGCTGGCCCTGCGAGGACAGCGGCGAGCGGTGAACATCTGTACCGAC 2160  
Qy 2161 AGCCAGTACCGCTTGGGCGATCATCCAGGCCCGCGACAGAGAGCGAGCGAGCTGGTG 2220  
Db 2161 AGCCAGTACCGCTTGGGCGATCATCCAGGCCCGCGACAGAGAGCGAGCGAGCTGGTG 2220  
Qy 2221 AACAGATCATTCAGCAGCTGATCAAGAGGAGAGGTGTACTTGTAGTGGTGGTGGTGGTGG 2280  
Db 2221 AACAGATCATTCAGCAGCTGATCAAGAGGAGAGGTGTACTTGTAGTGGTGGTGGTGGTGG 2280  
Qy 2281 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATCCGCAAG 2340  
Db 2281 CACAAGGCGATTCGCGCGCAACGAGCAGATCGAAGCTGTGTGAGCAAGGGGATCCGCAAG 2340  
Qy 2341 GTGCTGTTCCTGGAACGCGCATCGATGGGCGATCGTGTATCTACCAAGTACATGGACGACCTG 2400  
Db 2341 GTGCTGTTCCTGGAACGCGCATCGATGGGCGATCGTGTATCTACCAAGTACATGGACGACCTG 2400

QY 2401 TACGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
DB |||||  
2401 TACGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCACCGGTGAA 2460  
QY 2461 TTC 2463  
DB |||||  
2461 TTC 2463  
RESULT 2  
AX455914  
LOCUS AX455914 2469 bp DNA linear PAT 06-JUL-2002  
DEFINITION Sequence 30 from Patent WO20204493.  
ACCESSION AX455914  
VERSION AX455914.1 GI:21714899  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM synthetic construct  
other sequences; artificial sequences.  
1  
REFERENCE  
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.  
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,  
polypeptides and uses thereof  
JOURNAL Patent: WO 020493-A 30 17-JAN-2002;  
CHIRON CORPORATION (US); University of Stellenbosch (ZA)  
FEATURES  
source  
1..2469  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="PR975(+)"  
ORIGIN  
Query Match 99.2%; Score 2442.2; DB 6; Length 2469;  
Best Local Similarity 99.6%; Pred. No. 5.4e-246;  
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;  
QY 1 GTCGAGCCACCATGGCGAGGCCATGAGCCAGGCCACAGGCCCAACATCTGTATGAG 60  
DB 1 GTCGAGCCACCATGGCGAGGCCATGAGCCAGGCCACAGGCCCAACATCTGTATGAG 60  
QY 61 GCGAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGC 120  
DB 61 GCGAGCACTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGC 120  
QY 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGAGGGCTGTGAAAGTGGCGCAAGGAG 180  
DB 121 CACATCGCCGCAACTGCGCGCCCCCGCAAGAGAGGGCTGTGAAAGTGGCGCAAGGAG 180  
QY 181 GGCACACAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTCCGCGAGGACCTGGCC 240  
DB 181 GGCACACAGATGAAGGACTGCAACCGAGCGCCAGGCCAACTTCTCCGCGAGGACCTGGCC 240  
QY 241 TTCCCGCAGGCAAGCGCCGAGTTCCTCCAGCGAGCAACCGCGCCCAACAGCCCCACC 300  
DB 241 TTCCCGCAGGCAAGCGCCGAGTTCCTCCAGCGAGCAACCGCGCCCAACAGCCCCACC 300  
QY 301 AGCCCGAGCTGACAGTGGCGGCGCAACACCCCGCAGCGAGCGCGCGCGCGAGCGCCAG 360  
DB 301 AGCCCGAGCTGACAGTGGCGGCGCAACACCCCGCAGCGAGCGCGCGCGCGAGCGCCAG 360  
QY 361 GGCACCCCTGAATCTCCCGAGATCACTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG 420  
DB 361 GGCACCCCTGAATCTCCCGAGATCACTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG 420  
QY 421 GCGCGCCAGATCAAGAGGCCCTGTGGACACCGCGCGCGAGCACACCGTGTGTGAGGAG 480  
DB 421 GCGCGCCAGATCAAGAGGCCCTGTGTGGACACCGCGCGCGAGCACACCGTGTGTGAGGAG 480  
QY 481 ATGAGCTTCCCGCGCAAGTGGAGCCCAAGATGATGCGCGGCATCGCGGGCTTTCATCAAG 540  
DB 481 ATGAGCTTCCCGCGCAAGTGGAGCCCAAGATGATGCGCGGCATCGCGGGCTTTCATCAAG 540

QY 541 GTGCGCCAGTACGACAGATCTGTATCGAGATCTCGGGCAAGAGGCCATCGGACCCGTG 600  
DB 541 GTGCGCCAGTACGACAGATCTGTATCGAGATCTCGGGCAAGAGGCCATCGGACCCGTG 600  
QY 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
DB 601 CTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660  
QY 661 ACCCTGAATCTCCCATCAGCCCCATCAGACCCGTGCGCGTGAAGCTGAAGCCCGCATG 720  
DB 661 ACCCTGAATCTCCCATCAGCCCCATCAGACCCGTGCGCGTGAAGCTGAAGCCCGCATG 720  
QY 721 GACGCGCCCAAGGTGAAGCGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
DB 721 GACGCGCCCAAGGTGAAGCGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCC 780  
QY 781 ATCTCGGAGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
DB 781 ATCTCGGAGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
QY 841 AACACCCCGTGTTCGCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
DB 841 AACACCCCGTGTTCGCGCATCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900  
QY 901 TTCCGCGAGCTGAACAGGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGATCCCGCC 960  
DB 901 TTCCGCGAGCTGAACAGGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGATCCCGCC 960  
QY 961 CCGCGCGGCTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
DB 961 CCGCGCGGCTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
QY 1021 AGCGTGCCTTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
DB 1021 AGCGTGCCTTGGAGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGCATCAAC 1080  
QY 1081 AACGAGACCCCGGATCCGCTACAGTACAAAGTGTCTGCCAGGAGGTGGAAGGCGAGC 1140  
DB 1081 AACGAGACCCCGGATCCGCTACAGTACAAAGTGTCTGCCAGGAGGTGGAAGGCGAGC 1140  
QY 1141 CCCAGCATCTTCCAGAGCAGATGACCAAGATCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200  
DB 1141 CCCAGCATCTTCCAGAGCAGATGACCAAGATCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200  
QY 1201 GAGATCGTGTATCTACCA-----GGCGCCCTGTAGCTGGGCGAGCGACTGTGAGATCGGC 1254  
DB 1201 GAGATCGTGTATCTACCA-----GGCGCCCTGTAGCTGGGCGAGCGACTGTGAGATCGGC 1254  
QY 1255 CAGCACCGCGCAAGATCGAGAGGTGCGAAGCACTGTGCGGTGGGGCTTCAACGAC 1314  
DB 1255 CAGCACCGCGCAAGATCGAGAGGTGCGAAGCACTGTGCGGTGGGGCTTCAACGAC 1314  
QY 1315 CCCGACAGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1374  
DB 1315 CCCGACAGAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1374  
QY 1375 GACAAAGTGGAGCCGTGAGCCCATCGAGTGGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434  
DB 1375 GACAAAGTGGAGCCGTGAGCCCATCGAGTGGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434  
QY 1435 ATCCGAAAGTGTGGGCAAGCTGAATCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1494  
DB 1435 ATCCGAAAGTGTGGGCAAGCTGAATCTGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1494  
QY 1495 CGCGAGCTGTGCAAGTGTGCGGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1554  
DB 1495 CGCGAGCTGTGCAAGTGTGCGGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1554  
QY 1555 GAGGAGGCGGAGCTGGAGTGGCGGAGAAACCGCGAGATCTCTGCGGAGGAGGAGGAGGAGGAGGAG 1614  
DB 1555 GAGGAGGCGGAGCTGGAGTGGCGGAGAAACCGCGAGATCTCTGCGGAGGAGGAGGAGGAGGAGGAG 1614  
QY 1615 GTGTACTACGACCCCGAGCAAGGAGCTGTGTGGCGGAGATTCAGAGAGGAGGAGGAGGAGGAGGAG 1674  
DB 1615 GTGTACTACGACCCCGAGCAAGGAGCTGTGTGGCGGAGATTCAGAGAGGAGGAGGAGGAGGAGGAG 1674

```
|||||
Db 1621 GTGTACTACGACCCAGCAAGACCTGGTGGCCAGATCCAGAAGCAGGGCCACGACCAG 1680
Qy 1675 TGGACCTACCGATCTTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAG 1734
Db 1681 TGGACCTACCGATCTTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAG 1740
Qy 1735 ATGGCAGCCGCCACACCAAGAGTGAAGCAGCTGACCGAGGCGGTGCGAGAGATCGCC 1794
Db 1741 ATGGCAGCCGCCACACCAAGAGTGAAGCAGCTGACCGAGGCGGTGCGAGAGATCGCC 1800
Qy 1795 ATGAGAGATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTCCATCCAGAGGAG 1854
Db 1801 ATGAGAGATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTTCCGCTCCATCCAGAGGAG 1860
Qy 1855 ACCTGGAGACCTGGTGGACCGCACTATGCGAGCGCCACCTGGATCCCGAGTGGGAGTTC 1914
Db 1861 ACCTGGAGACCTGGTGGACCGCACTATGCGAGCGCCACCTGGATCCCGAGTGGGAGTTC 1920
Qy 1915 GTGAACACCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGC 1974
Db 1921 GTGAACACCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGGCCATCATCGGC 1980
Qy 1975 GCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGSCCGGC 2034
Db 1981 GCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGSCCGGC 2040
Qy 2035 TACGTGACCGACCGGGCGCGCAGAGATCTGTAGCTGACCGAGACCAACCAAGAG 2094
Db 2041 TACGTGACCGACCGGGCGCGCAGAGATCTGTAGCTGACCGAGACCAACCAAGAG 2100
Qy 2095 ACCGAGCTGGAGGCCATCCAGCTGGCCCTGCGAGCAGCGCAGCGAGGTGAACATCGTG 2154
Db 2101 ACCGAGCTGGAGGCCATCCAGCTGGCCCTGCGAGCAGCGCAGCGAGGTGAACATCGTG 2160
Qy 2155 ACCGAGCAGCAGTAGTACGGCCCTGGGCGATCATCAGGCCCGACCGCCGACGAGCGAG 2214
Db 2161 ACCGAGCAGCAGTAGTACGGCCCTGGGCGATCATCAGGCCCGACCGCCGACGAGCGAG 2220
Qy 2215 CTGTGTAACCGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2274
Db 2221 CTGTGTAACCGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2280
Qy 2275 CCGCCACACAGGCGATCGCGGCAAGCAGCAGATCGACAGCTGTGAGCAAGGGCATC 2334
Db 2281 CCGCCACACAGGCGATCGCGGCAAGCAGCAGATCGACAGCTGTGAGCAAGGGCATC 2340
Qy 2335 CGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGGAC 2394
Db 2341 CGCAAGGTGCTGTTCTTGGACGGCATCGATGGCGGCATCGTGATCTACCAAGTACATGGAC 2400
Qy 2395 GACCTGATCGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2454
Db 2401 GACCTGATCGTGGCAGCGCGCCCTAGGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
Qy 2455 GGTGAATTC 2463
Db 2461 GGTGAATTC 2469
```

```
RESULT 3
AX455916
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS :
SOURCE :
ORGANISM :
REFERENCE
1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
Polynucleotides encoding antigenic hiv type c polypeptides,
```

```
polypeptides and uses thereof
Patent: WO 0204493-A 32 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source Location/Qualifiers
1. .2457
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YMMW"
ORIGIN
Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
Best Local Similarity 99.6%; Pred. No. 2.3e-245;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
Qy 1 GTGAGCCCAACCTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 60
Db 1 GTGAGCCCAACCTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 60
Qy 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120
Db 61 CGCAGCAACTTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGC 120
Qy 121 CACATCGCCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180
Db 121 CACATCGCCCGCAACTGCGCGGCCCGCCGCAAGAGGGCTGTGGAAGTGGCGCAAGGAG 180
Qy 181 GGCCACAGATGAAGGATGCAACGAGCGCCAGCCCAACTTCTCCGAGGAGCCTGGCC 240
Db 181 GGCCACAGATGAAGGATGCAACGAGCGCCAGCCCAACTTCTCCGAGGAGCCTGGCC 240
Qy 241 TTCCCCCAGGCGAAGGCCCGCGAGTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCCACC 300
Db 241 TTCCCCCAGGCGAAGGCCCGCGAGTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCCACC 300
Qy 301 AGCGCGAGCTGCAAGTGTGCGCGGCAACACCCCGCAGCGAGGCCGCGCGCGAGCGCGAG 360
Db 301 AGCGCGAGCTGCAAGTGTGCGCGGCAACACCCCGCAGCGAGGCCGCGCGCGAGCGCGAG 360
Qy 361 GGCACCTTGAACCTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG 420
Db 361 GGCACCTTGAACCTTCCCGCAGATCACCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTG 420
Qy 421 GCGCGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCCCGACGACACCGTCTGGAGGAG 480
Db 421 GCGCGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCCCGACGACACCGTCTGGAGGAG 480
Qy 481 ATGAGCTGCGCCCGCAAGTGGAAAGCCCAAGATGATGCGCGGCATGCGCGGCTTCATCAAG 540
Db 481 ATGAGCTGCGCCCGCAAGTGGAAAGCCCAAGATGATGCGCGGCATGCGCGGCTTCATCAAG 540
Qy 541 GTGCGCCAGTACGACAGATCTTGATCGAGATCTGCGGCAAGGAGGCCATCGGCACCGTG 600
Db 541 GTGCGCCAGTACGACAGATCTTGATCGAGATCTGCGGCAAGGAGGCCATCGGCACCGTG 600
Qy 601 CTGATCGCGCCACCCCGCGTGAACATCATCGCGCGCAACATGTCTGACCCAGCTGGGTGC 660
Db 601 CTGATCGCGCCACCCCGCGTGAACATCATCGCGCGCAACATGTCTGACCCAGCTGGGTGC 660
Qy 661 ACCCTGAACTTCCCATCAGCCCCCATCGAGACCGCTGCGCGCAAGAGGCCATCGGCACCGTG 720
Db 661 ACCCTGAACTTCCCATCAGCCCCCATCGAGACCGCTGCGCGCAAGAGGCCATCGGCACCGTG 720
Qy 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC 780
Db 721 GACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAGATCAAGGGCCCTGACCGCC 780
Qy 781 ATCTGCGAGGAGATGGAGAGGAGGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTAC 840
Db 781 ATCTGCGAGGAGATGGAGAGGAGGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTAC 840
Qy 841 AACACCCCGTGTTCATCAAGAGAGAGGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTAC 900
Db 841 AACACCCCGTGTTCATCAAGAGAGAGGAGGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTAC 900
```



841 AACACCCCGGTGTTCCGCATCAAGAGAGAGACAGCAACCAAGTGGCGCAAGCTGGTGGAC 900  
901 TTCCCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCATCCCCCAG 960  
901 TTCCCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGCACTGGGCATCCCCCAG 960  
961 CCCCGCGGCTGAAGAGAGAGAGCTGACCGTCTGGAGGTGCACTGGGCATCCCCCAG 1020  
961 CCCCGCGGCTGAAGAGAGAGAGCTGACCGTCTGGAGGTGCACTGGGCATCCCCCAG 1020  
1021 AGCGTGCCTCCGAGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
1021 AGCGTGCCTCCGAGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAAC 1080  
1081 AACGAGACCCCGGATCCGCTACCAAGTACCAAGTCTGCTGGAGGTGCACTGGGCATCCCCCAG 1140  
1081 AACGAGACCCCGGATCCGCTACCAAGTACCAAGTCTGCTGGAGGTGCACTGGGCATCCCCCAG 1140  
1141 CCCAGCATCTCCAGAGCAGATGACCAAGATCTCTGGAGGTGCACTGGGCATCCCCCAG 1200  
1141 CCCAGCATCTCCAGAGCAGATGACCAAGATCTCTGGAGGTGCACTGGGCATCCCCCAG 1200  
1201 GAGATCGTGTATACACAGCCCGCTGTACGTGGGCGAGCGCTGGAGATCGGCCAGCAC 1260  
1201 GAGATCGTGTATACACAGCCCGCTGTACGTGGGCGAGCGCTGGAGATCGGCCAGCAC 1260  
1261 CGCGCAAGATCGAGAGCTGGCGAAGCACCTGTGCGCTGGGCTTCAACACCCCGGAC 1320  
1261 CGCGCAAGATCGAGAGCTGGCGAAGCACCTGTGCGCTGGGCTTCAACACCCCGGAC 1320  
1321 AAGAGACACAG 1380  
1321 AAGAGACACAG 1380  
1375 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1375 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
1441 AAGCTGTGGGCAAGCTGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
1441 AAGCTGTGGGCAAGCTGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
1435 AAGCTGTGGGCAAGCTGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494  
1435 AAGCTGTGGGCAAGCTGAACTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494  
1501 CTGTGCAAGCTGTGCGGCGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
1501 CTGTGCAAGCTGTGCGGCGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560  
1495 CTGTGCAAGCTGTGCGGCGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554  
1495 CTGTGCAAGCTGTGCGGCGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554  
1561 GCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGAGAGAGAGAGAGAGAG 1620  
1561 GCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGAGAGAGAGAGAGAGAG 1620  
1555 GCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGAGAGAGAGAGAGAGAG 1614  
1555 GCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGAGAGAGAGAGAGAGAG 1614  
1621 TACGACCCAGCAGAGACTGTGTGGCGGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1621 TACGACCCAGCAGAGACTGTGTGGCGGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAG 1680  
1615 TACGACCCAGCAGAGACTGTGTGGCGGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAG 1674  
1615 TACGACCCAGCAGAGACTGTGTGGCGGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAG 1674  
1681 TACGAGATCTACAG 1740  
1681 TACGAGATCTACAG 1740  
1675 TACGAGATCTACAG 1734  
1675 TACGAGATCTACAG 1734  
1741 ACCGCCACACAG 1800  
1741 ACCGCCACACAG 1800  
1735 ACCGCCACACAG 1794  
1735 ACCGCCACACAG 1794  
1801 AGCATCGTGTATCTGGGCGAG 1860  
1801 AGCATCGTGTATCTGGGCGAG 1860  
1795 AGCATCGTGTATCTGGGCGAG 1854  
1795 AGCATCGTGTATCTGGGCGAG 1854  
1861 GAGACCTGGTGGAG 1920  
1861 GAGACCTGGTGGAG 1920  
1855 GAGACCTGGTGGAG 1914  
1855 GAGACCTGGTGGAG 1914  
1921 ACCGCCACACAG 1980  
1921 ACCGCCACACAG 1980  
1915 ACCGCCACACAG 1974  
1915 ACCGCCACACAG 1974

1981 ACCTTTCTAGTGGACGGCGCGCCCAACCGCGAGAGCAAGAGATCGGCAAGCGCGCTACGTG 2040  
1981 ACCTTTCTAGTGGACGGCGCGCCCAACCGCGAGAGCAAGAGATCGGCAAGCGCGCTACGTG 2040  
1975 ACCTTTCTAGTGGACGGCGCGCCCAACCGCGAGAGCAAGAGATCGGCAAGCGCGCTACGTG 2034  
1975 ACCTTTCTAGTGGACGGCGCGCCCAACCGCGAGAGCAAGAGATCGGCAAGCGCGCTACGTG 2034  
2041 ACCGACCGGGCGCGCAGAGAGATCGTGGAGCTGACCGGAGAGACCAACCAAGAGACCGGAG 2100  
2041 ACCGACCGGGCGCGCAGAGAGATCGTGGAGCTGACCGGAGAGACCAACCAAGAGACCGGAG 2100  
2035 ACCGACCGGGCGCGCAGAGAGATCGTGGAGCTGACCGGAGAGACCAACCAAGAGACCGGAG 2094  
2035 ACCGACCGGGCGCGCAGAGAGATCGTGGAGCTGACCGGAGAGACCAACCAAGAGACCGGAG 2094  
2101 CTGACGGCCATCCAGCTGGCCCTCGAGGACAGCGGACGAGGTTGAACATCTGTGACCGGAC 2160  
2101 CTGACGGCCATCCAGCTGGCCCTCGAGGACAGCGGACGAGGTTGAACATCTGTGACCGGAC 2160  
2095 CTGACGGCCATCCAGCTGGCCCTCGAGGACAGCGGACGAGGTTGAACATCTGTGACCGGAC 2154  
2095 CTGACGGCCATCCAGCTGGCCCTCGAGGACAGCGGACGAGGTTGAACATCTGTGACCGGAC 2154  
2161 AGCAGTACGCTTGGGATCATTCAGGCCCGAGCCCGCAAGAGAGAGAGGAGTGTGTG 2220  
2161 AGCAGTACGCTTGGGATCATTCAGGCCCGAGCCCGCAAGAGAGAGAGGAGTGTGTG 2220  
2155 AGCAGTACGCTTGGGATCATTCAGGCCCGAGCCCGCAAGAGAGAGAGGAGTGTGTG 2214  
2155 AGCAGTACGCTTGGGATCATTCAGGCCCGAGCCCGCAAGAGAGAGAGGAGTGTGTG 2214  
2221 AACCAGATCATCGAGGAGCTGATCAAGAGAGAGAGGAGTGTACCTGAGCTGGTGGTCCGCC 2280  
2221 AACCAGATCATCGAGGAGCTGATCAAGAGAGAGAGGAGTGTACCTGAGCTGGTGGTCCGCC 2280  
2215 AACCAGATCATCGAGGAGCTGATCAAGAGAGAGAGGAGTGTACCTGAGCTGGTGGTCCGCC 2274  
2215 AACCAGATCATCGAGGAGCTGATCAAGAGAGAGAGGAGTGTACCTGAGCTGGTGGTCCGCC 2274  
2281 CACAAGGCGATCGGCGGCAAGCAGCAGATCGAAGCTGGTGGAGCAAGGCGATCCGCAAG 2340  
2281 CACAAGGCGATCGGCGGCAAGCAGCAGATCGAAGCTGGTGGAGCAAGGCGATCCGCAAG 2340  
2275 CACAAGGCGATCGGCGGCAAGCAGCAGATCGAAGCTGGTGGAGCAAGGCGATCCGCAAG 2334  
2275 CACAAGGCGATCGGCGGCAAGCAGCAGATCGAAGCTGGTGGAGCAAGGCGATCCGCAAG 2334  
2341 GTGCTGTTCTTGGAGCGGATCGATGGCGGATCTGTGATCTACCATGATCATGAGACGCTG 2400  
2341 GTGCTGTTCTTGGAGCGGATCGATGGCGGATCTGTGATCTACCATGATCATGAGACGCTG 2400  
2335 GTGCTGTTCTTGGAGCGGATCGATGGCGGATCTGTGATCTACCATGATCATGAGACGCTG 2394  
2335 GTGCTGTTCTTGGAGCGGATCGATGGCGGATCTGTGATCTACCATGATCATGAGACGCTG 2394  
2401 TACGTGGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2460  
2401 TACGTGGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2460  
2395 TACGTGGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2454  
2395 TACGTGGGCGAGCGGCGGCTTAGGATCGATTAAAGCTTCCGGGGCTAGCACCGGTGAA 2454  
2461 TTC 2463  
2461 TTC 2463  
2455 TTC 2457  
2455 TTC 2457

RESULT 4  
BD263704  
LOCUS  
DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.  
ACCESSION BD263704  
VERSION BD263704.1 GI:33073472  
KEYWORDS JP 2002533124-A/71.  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1 (bases 1 to 2306)  
AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: JP 2002533124-A 71 08-OCT-2002;  
COMMENT CHIRON CORP  
OS Artificial Sequence  
PN JP 2002533124-A/71  
PD 08-OCT-2002  
PF 30-DEC-1999 JP 2000591193  
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI  
SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI KARIN HARTOG,  
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC  
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC  
C12N5/10,  
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC  
Description of Artificial Sequence: FS(-).protmod.Rtopt.YM FH Key  
Location/Qualifiers  
FT source 1..2306  
FT /organism='Artificial Sequence'.  
Location/Qualifiers  
1..2306

		/organism="synthetic construct"					
		/mol_type="genomic DNA"					
		/db_xref="taxon:32630"					
ORIGIN							
Query Match		83.1%; Score 2046; DB 6; Length 2306;					
Best Local Similarity		93.6%; Pred. No. 1.1e-204;					
Matches 2159; Conservative		0; Mismatches 135; Indels 12; Gaps 2;					
Qy	170	GC	GC	GC	GC	GC	GC
Db	1	GC	GC	GC	GC	GC	GC
Qy	230	AG	AG	AG	AG	AG	AG
Db	61	AG	AG	AG	AG	AG	AG
Qy	290	AC	AC	AC	AC	AC	AC
Db	121	AC	AC	AC	AC	AC	AC
Qy	344	CC	CC	CC	CC	CC	CC
Db	181	CC	CC	CC	CC	CC	CC
Qy	398	GC	GC	GC	GC	GC	GC
Db	241	GC	GC	GC	GC	GC	GC
Qy	458	CC	CC	CC	CC	CC	CC
Db	301	CC	CC	CC	CC	CC	CC
Qy	518	GC	GC	GC	GC	GC	GC
Db	361	GC	GC	GC	GC	GC	GC
Qy	578	GC	GC	GC	GC	GC	GC
Db	421	GC	GC	GC	GC	GC	GC
Qy	638	AC	AC	AC	AC	AC	AC
Db	481	AC	AC	AC	AC	AC	AC
Qy	698	CC	CC	CC	CC	CC	CC
Db	541	CC	CC	CC	CC	CC	CC
Qy	758	AG	AG	AG	AG	AG	AG
Db	601	AG	AG	AG	AG	AG	AG
Qy	818	AG	AG	AG	AG	AG	AG
Db	661	AG	AG	AG	AG	AG	AG
Qy	878	CC	CC	CC	CC	CC	CC
Db	721	CC	CC	CC	CC	CC	CC
Qy	938	AG	AG	AG	AG	AG	AG
Db	781	AG	AG	AG	AG	AG	AG
Qy	998	TG	TG	TG	TG	TG	TG
Db	841	TG	TG	TG	TG	TG	TG
Qy	1058	CC	CC	CC	CC	CC	CC
Db	901	CC	CC	CC	CC	CC	CC
Qy	1118	TG	TG	TG	TG	TG	TG
Db	961	TG	TG	TG	TG	TG	TG
Qy	1178	AG	AG	AG	AG	AG	AG
Db	1021	AG	AG	AG	AG	AG	AG
Qy	1238	GC	GC	GC	GC	GC	GC
Db	1081	GC	GC	GC	GC	GC	GC
Qy	1298	GT	GT	GT	GT	GT	GT
Db	1141	GT	GT	GT	GT	GT	GT
Qy	1358	GT	GT	GT	GT	GT	GT
Db	1201	GT	GT	GT	GT	GT	GT
Qy	1418	GT	GT	GT	GT	GT	GT
Db	1261	GT	GT	GT	GT	GT	GT
Qy	1478	AC	AC	AC	AC	AC	AC
Db	1321	AC	AC	AC	AC	AC	AC
Qy	1538	AC	AC	AC	AC	AC	AC
Db	1381	AG	AG	AG	AG	AG	AG
Qy	1598	GC	GC	GC	GC	GC	GC
Db	1441	AG	AG	AG	AG	AG	AG
Qy	1658	AG	AG	AG	AG	AG	AG
Db	1501	AG	AG	AG	AG	AG	AG
Qy	1718	CC	CC	CC	CC	CC	CC
Db	1561	CC	CC	CC	CC	CC	CC
Qy	1778	CC	CC	CC	CC	CC	CC
Db	1621	CC	CC	CC	CC	CC	CC
Qy	1838	TG	TG	TG	TG	TG	TG
Db	1681	TG	TG	TG	TG	TG	TG
Qy	1898	TC	TC	TC	TC	TC	TC
Db	1741	TC	TC	TC	TC	TC	TC
Qy	1958	AG	AG	AG	AG	AG	AG
Db	1801	AG	AG	AG	AG	AG	AG
Qy	2018	AG	AG	AG	AG	AG	AG
Db	1861	AG	AG	AG	AG	AG	AG
Qy	2078	AG	AG	AG	AG	AG	AG
Db	1921	AC	AC	AC	AC	AC	AC
Qy	2138	GC	GC	GC	GC	GC	GC
Db	1981	TG	TG	TG	TG	TG	TG
Qy	2198	AC	AC	AC	AC	AC	AC

Db 2041 ACAAGAGCCAGACGAGCTGGTGAGCCAGATCATCGAGCGCTGATCAAGAGGAGAGG 2100  
Qy 2258 TGTACTGTGAGCTGGGTGCGCCGCCACAGGGGATCGCGGCAACGAGCAGATCGACAAGC 2317  
Db 2101 TGTACTGTGCGCTGGGTGCGCCGCCACAGGGGATCGCGGCAACGAGCAGATCGACAAGC 2160  
Qy 2318 TGGTGAAGCAGGCGATCCCGAAGGTGCTGTTCTTGAAGCGGATCGATGCGCGGATCGTGA 2377  
Db 2161 TGGTGAAGCAGGCGATCCCGAAGGTGCTGTTCTTGAAGCGGATCGATGCGCGGATCGTGA 2220  
Qy 2378 TCTACAGTACATGACGACCTGTAGTGGGAGCGGCGGCGCTAGGATCGATTAAAGC 2437  
Db 2221 TCTACAGTACATGACGACCTGTAGTGGGAGCGGCGGCGCTAGGATCGATTAAAGC 2280  
Qy 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463  
Db 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 5  
LOCUS Q0870574 2306 bp DNA linear PAT 13-SEP-2004  
DEFINITION Sequence 82 from Patent EP1433851.  
ACCESSION Q0870574  
VERSION Q0870574.1 GI:52000090  
KEYWORDS  
SOURCE  
ORGANISM  
synthetic construct  
other sequences; artificial sequences.

REFERENCE 1  
AUTHORS Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K.,  
Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: EP 1433851-A 82 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES  
source  
1..2306  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Description of Artificial Sequence:  
FS(-).protmod.RTopt.YM"

ORIGIN  
Query Match 83.1%; Score 2046; DB 6; Length 2306;  
Best Local Similarity 93.6%; Pred. No. 1.1e-204;  
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
Qy 170 GCGCAAGGAGGCGCCACAGATGAAGGACTGCACCGAGCGCCAGGCGCAACTTTCTTCGCG 229  
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCCTGAGAGACAGGCTAATTCTTTCGCG 60  
Qy 230 AGGACTGGCTTCCCGCAGGCAAGCGCGGAGTTCGCCAGCAGCAGCAACCGCGCA 289  
Db 61 AGGACTGGCTTCTGCGGCGCAAGCGCGGAGTTCAGCAGCAGCAGCAACCGCGCA 120  
Qy 290 ACAGCCCGCAGCGCGGAGTTCAGGTCGCGGCG-----ACAACCCCGCAGCGAGG 343  
Db 121 ACAGCCCGCAGCGCGGAGTTCAGGTCGCGGCG-----ACAACCCCGCAGCGAGG 180  
Qy 344 CCGGCGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 397  
Db 181 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240  
Qy 398 GCCCTCTGTGAGCATCAAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 457  
Db 241 GCCCTCTGTGAGCATCAAGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
Qy 458 CCGAGCACCGTGTGAGGAGATGAGCTGCCCGGCAAGTGGAGGCGGCGGCGGCGGCGGCGGCGG 517  
Db 301 CCGAGCACCGTGTGAGGAGATGAGCTGCCCGGCAAGTGGAGGCGGCGGCGGCGGCGGCGGCGG 360

Qy 518 GCGGATCGGCGGCTTTCATCAAGGTGCGCGCAGTACGACAGATCCTGTGATCGAGATCTGG 577  
Db 361 GCGGATCGGCGGCTTTCATCAAGGTGCGCGCAGTACGACAGATCCTGTGATCGAGATCTGG 420  
Qy 578 GCAGAAGCCATCGGCGACCGTGTGATCGGCGCCACCGCGTGAACATCATCGGCGCA 637  
Db 421 GCCACAAGCCATCGGCGACCGTGTGATCGGCGCCACCGCGTGAACATCATCGGCGCA 480  
Qy 638 ACATCTGACCCAGCTGGGCTGCACCTCGAACTTCCCATCAGCGCCCATCGAGACCGTGC 697  
Db 481 ACCTGCTGACCCAGATCGGCTGCACCTCGAACTTCCCATCAGCGCCCATCGAGACCGTGC 540  
Qy 698 CCGTGAAGCTGAAGCCCGGCGATGAGCGGCGCCCAAGGTGAAGCAGTGGCCCTTCACCGAGG 757  
Db 541 CCGTGAAGCTGAAGCCCGGCGATGAGCGGCGCCCAAGGTCAAGCAGTGGCCCTTCACCGAGG 600  
Qy 758 AGAGATCAAGGCCCTGACCGCATCTCGAGGAGATGAGAGAGGAGGCGCAAGATCACCA 817  
Db 601 AGAGATCAAGGCCCTGAGGATCTGACCGAGATGAGAGAGGAGGCGCAAGATCACCA 660  
Qy 818 AGATCGGCGCGGAGAACCCCTACACACCCCGCTGTTCCCATCAAGAGAGAGCAGCA 877  
Db 661 AGATCGGCGCGGAGAACCCCTACACACCCCGCTGTTCCCATCAAGAGAGAGCAGCA 720  
Qy 878 CCAAGTGGCGCAAGCTGTGTGACTTCCCGGAGCTGAACAAGCGCAACCAGGACTTCTGGG 937  
Db 721 CCAAGTGGCGCAAGCTGTGTGACTTCCCGGAGCTGAACAAGCGCAACCAGGACTTCTGGG 780  
Qy 938 AGGTGACGTGGGATCCCGCCCGCGGCTGAAGAGAGAGAGCGGTGACCGTGC 997  
Db 781 AGGTGACGTGGGATCCCGCCCGCGGCTGAAGAGAGAGAGCGGTGACCGTGC 840  
Qy 998 TGGAGTGGGCGAAGCTTACTTTCAGCGTGCCTTCGAGCAGGACTTCCCGCAAGTACACCG 1057  
Db 841 TGGAGTGGGCGAAGCTTACTTTCAGCGTGCCTTCGAGCAGGACTTCCCGCAAGTACACCG 900  
Qy 1058 CTTTACCATCCCGAGCATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117  
Db 901 CTTTACCATCCCGAGCATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Qy 1118 TGCCCCAGGCTGGAAGGCGAGCGCCAGCATCTTCCAGAGCAGCATGACCCAGAGATCTGG 1177  
Db 961 TGCCCCAGGCTGGAAGGCGAGCGCCAGCATCTTCCAGAGCAGCATGACCCAGAGATCTGG 1020  
Qy 1178 AGCCCTTCGCGCGCGCAACCCCGAGATCTGATCTACAGCGCCCGCTGTACGTGGCA 1237  
Db 1021 AGCCCTTCGCGCGAG 1080  
Qy 1238 GCGACCTGGAGATCGGCGAGCAACCGCGCAAGATCGAGAGGTGGCGAAGCAGCTGTGC 1297  
Db 1081 GCGACCTGGAGATCGGCGAGCAACCGCGCAAGATCGAGAGGTGGCGGAGCTGTGCTGTGC 1140  
Qy 1298 GCTGGGCTTCAACACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357  
Db 1141 GCTGGGCTTCAACACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
Qy 1358 GCTAGAGCTGCAACCCCGCAAGTGGAGCGGTGACCGCATCGAGCTGCCGAGAGAGAGAG 1417  
Db 1201 GCTAGAGCTGCAACCCCGCAAGTGGAGCGGTGACCGCATCATGCTGCCGAGAGAGAGAG 1260  
Qy 1418 GCTGAGCGGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGCAGAGATCT 1477  
Db 1261 GCTGAGCGGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGCAGAGATCT 1320  
Qy 1478 ACCCGGATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGGCGGCGCAGAGCGCTGACCG 1537  
Db 1321 ACAGCGGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGGCGGCGCAGAGCGCTGACCG 1380  
Qy 1538 ACATCTGCGCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAGAGAGAGAGAGAGAGAGAG 1597  
Db 1381 AGGTATCCCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAGAGAGAGAGAGAGAGAGAG 1440

Qy 1598 GCGAGCCGTGACGCGGTGTACTACGACCCAGAGGACCTGGTGGCGAGATCCAGA 1657  
Db |||||||  
Qy 1441 AGGAGCCGTGCACGAGGTGTACTACGACCCACAGGACCTGGTGGCGAGATCCAGA 1500  
Db |||||||  
Qy 1658 AGCAGGCCACGACGACGAGTGGACCTTACCAGATCTACGAGGAGCCCTTCAAGAACCTGAAGA 1717  
Db |||||||  
Qy 1501 AGCAGGCCACGAGGCCAGTGGACCTTACCAGATCTACGAGGAGCCCTTCAAGAACCTGAAGA 1560  
Db |||||||  
Qy 1718 CCGGCAAGTACGCAAGATGGCCACCGCCACACCAAGAGCTGAAGAGTGCACGAGG 1777  
Db |||||||  
Qy 1561 CCGGCAAGTACGCCCGATGGCGCGCCACACCAACGACGTAAGCAGCTGACCGAGG 1620  
Db |||||||  
Qy 1778 CCGTGCAGAAAGTCCGATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGC 1837  
Db |||||||  
Qy 1621 CCGTGCAGAAAGTGCAGCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1680  
Db |||||||  
Qy 1838 TGCCCATCCAGAAAGGAGACCTTGGGAGACCTGGTGGACCGACTACTTGGCAGGCCACCTGGA 1897  
Db |||||||  
Qy 1681 TGCCCATCCAGAAAGGAGACCTTGGGAGGCTTGGTGGATGGAGTACTTGGCAGGCCACCTGGA 1740  
Db |||||||  
Qy 1898 TCCCGAGTGGGAGTTCGTGAAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGAGA 1957  
Db |||||||  
Qy 1741 TCCCGAGTGGGAGTTCGTGAAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGAGA 1800  
Db |||||||  
Qy 1958 AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCA 2017  
Db |||||||  
Qy 1801 AGGAGCCCATCGTGGGCGCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCA 1860  
Db |||||||  
Qy 2018 AGATCGGCAAGCCGGCTTACGTGACCGACCGGGGCGCGAGAGATCGTGAAGCTGACCG 2077  
Db |||||||  
Qy 1861 AGCTGGGCAAGCCGGCTACGTGACCGACCGGGCGCGGAGAGGTGGTGAAGCATCGCG 1920  
Db |||||||  
Qy 2078 AGACCAACCAAGAGACCGAGCTGCAGGCCATTCAGCTGGCCCTGGAGAGACGAGCGCA 2137  
Db |||||||  
Qy 1921 ACACCAACCAAGAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCGAGGACAGCGCC 1980  
Db |||||||  
Qy 2138 GCGAGGTGAACATCGTGACCGACAGCAGCAGTACGCCCTGGGATCATTCAGGCCCGAGCCG 2197  
Db |||||||  
Qy 1981 TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGATCATTCAGGCCCGAGCCG 2040  
Db |||||||  
Qy 2198 ACAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGGAAG 2257  
Db |||||||  
Qy 2041 ACAAGAGCGAGCGAGCTGGTGAAGCCAGATCATCGAGCAGCTGATCAAGAGGAGGAAG 2100  
Db |||||||  
Qy 2258 TGTACTGAGCTGGGTGCGCCGCCACAGGCGCATCGGGCGCAACGAGCAGATCCAGCAGC 2317  
Db |||||||  
Qy 2101 TGTACTGGCTGGGTGCGCCGCCACAGGCGCATCGGGCGCAACGAGCAGGTGACCAAGC 2160  
Db |||||||  
Qy 2318 TGGTGAGCAAGGCATCCGCAAGTGTCTTCTGGACGGCATCGATGGCGGCGATCGTGA 2377  
Db |||||||  
Qy 2161 TGGTGAGCGCGGCATCCGCAAGTGTCTTCTGAAACGGCATCGATGGCGGCGATCGTGA 2220  
Db |||||||  
Qy 2378 TCTACCAAGTACATGGAACGACCTGTACGTGGGCGAGCGCGCCCTAGGATCGATTAAAGC 2437  
Db |||||||  
Qy 2221 TCTACCAAGTACATGGAACGACCTGTACGTGGGCGAGCGCGCCCTAGGATCGATTAAAGC 2280  
Db |||||||  
Qy 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463  
Db |||||||  
Qy 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306  
Db |||||||

RESULT 6  
AR373387  
LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 82 from patent US 6602705.  
ACCESSION AR373387  
VERSION AR373387.1 GI:40075490  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2306)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.

TITLE Expression of HIV polypeptides and production of virus-like particles  
JOURNAL Patent: US 6602705-A 82 05-AUG-2003;  
FEATURES Location/Qualifiers  
source 1. .2306  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 83.1%; Score 2046; DB 6; Length 2306;  
Best Local Similarity 93.6%; Pred. No. 1.1e-204;  
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
Qy 170 GCGCAAGGAGGGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCGC 229  
Db |||||||  
Qy 230 AGGACCTTGGCTTCCCGGAGGCAAGCCGCGAGTTCCCGAGCGAGCAGAACCGCGCCA 289  
Db |||||||  
Qy 61 AGGACCTTGGCTTCCCGAGGCAAGCCGCGAGTTTCAGCGAGCAGACCGCGGCCA 120  
Db |||||||  
Qy 290 ACAGCCCCACAGCGCGGAGCTGCAGGTGCGCGCGC-----ACAAACCCCGCAGCGAGG 343  
Db |||||||  
Qy 121 ACAGCCCCACCGCGCGGAGCTGCAGGTGTGGGGCGCGGAGAACAAAGCGCTGAGCGAGG 180  
Db |||||||  
Qy 344 CCGCGCGCGGAGCGCGAGGCAACCTCG-----AACTTCCCGCAGATCACTCTGTGGCAGC 397  
Db |||||||  
Qy 181 CCGCGCGCGGAGCGCGAGGCAACCTGAGCTTCAACTTCCCGCAGATCACTCTGTGGCAGC 240  
Db |||||||  
Qy 398 GCCCGCTGTGTGAGCATCAAGTGTGGCGCGCAGATCAAGGAGGCCCTCTGTGACACCGCGC 457  
Db |||||||  
Qy 241 GCCCGCTGTGTGAGCATCAAGTGTGGCGCGCAGCTCAAGGAGGCCCTCTGTGACACCGCGC 300  
Db |||||||  
Qy 458 CCGACGACACCGCTGTGTGAGGAGATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATCG 517  
Db |||||||  
Qy 301 CCGACGACACCGCTGTGTGAGGAGATGAACCTTCCCGCGCAAGTGAAGCCCAAGATGATCG 360  
Db |||||||  
Qy 518 GCGGCATCGCGCGCTTCATCAAGGTGCGCCAGTACGACACAGATCCTCTGATCGAGATCTGCG 577  
Db |||||||  
Qy 361 CCGGATCGGGGGCTTCATCAAGTGTGGCAGTACGACAGATCCCGCTGGAGATCTGCG 420  
Db |||||||  
Qy 578 GCAAGAGGCGCATCGGCAACCGTGTGATCGCGCCCAACCGCGTGAATCATCATCGCGCGCA 637  
Db |||||||  
Qy 421 GCCACAAAGGCGCATCGGCAACCGTGTGTGGGGCGCCACCGCGTGAACATCATCGCGCGCA 480  
Db |||||||  
Qy 638 ACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCGCATCAGCCCATCGAGACCGTGC 697  
Db |||||||  
Qy 481 ACCTGCTGACCCAGATCGGCTGCACCTGAACTTCCCGCATCAGCCCATCGAGACCGGTGC 540  
Db |||||||  
Qy 698 CCGTGAAGCTGAAGCCCGCATGGAACCGCCCAAGGTGAAGCAGTGGCCCCCTGACCCGAGG 757  
Db |||||||  
Qy 541 CCGTGAAGCTGAAGCCCGGAGTGAACCGCCCAAGGTCAAGCAGTGGCCCCCTGACCCGAGG 600  
Db |||||||  
Qy 758 AGAAGATCAAGGCGCTTGAACCGCATCTCGAGGAGATGGAAGAGGAGGCAAGATCAACA 817  
Db |||||||  
Qy 601 AGAAGATCAAGGCGCTTGTGTGAGATCTGCACCGAGATGGAAGAGGAGGCAAGATCAACA 660  
Db |||||||  
Qy 818 AGATCGGCGCGGAGAACCCCTTACAACACCCCGCTGTTCGCCCATCAAGAGAGGAGCAGCA 877  
Db |||||||  
Qy 661 AGATCGGCGCGGAGAACCCCTTACAACACCCCGCTGTTCGCCCATCAAGAGAGGAGCAGCA 720  
Db |||||||  
Qy 878 CCAAGTGGCGCAAGCTGTGTGACTTCCCGAGCTTGAACNAGCGCAACCGAGCTTCTGGG 937  
Db |||||||  
Qy 721 CCAAGTGGCGCAAGCTGTGTGACTTCCCGAGCTTGAACNAGCGCAACCGAGCTTCTGGG 780  
Db |||||||  
Qy 938 AGGTGACAGCTGGGCGATCCCCACCCCGCGCTTGAAGAGAGGAGGAGGCGTGAACCGTGC 997  
Db |||||||  
Qy 781 AGGTGACAGCTGGGCGATCCCCCACCCCGCGCTTGAAGAGAGGAGGAGGCGTGAACCGTGC 840  
Db |||||||  
Qy 998 TGGACGTGGGCGAGCGCTTACTTTCAGCTGCGCCCTGGAGGAGGACTTCCGCAAGTACACCG 1057  
Db |||||||  
Qy 841 TGGACGTGGGCGAGCGCTTACTTTCAGCTGCGCCCTGGAGCAAGGACTTCCGCAAGTACACCG 900  
Db |||||||

1058 QY CTTTCACCATCCCGAGCATCAACAGAGAGACCCCGGCATCGGTACCAAGTACACGTCG 1117  
901 Db CTTTCACCATCCCGAGCATCAACAGAGAGACCCCGGCATCGGTACCAAGTACACGTCG 960  
1118 QY TGCCCCAGGGCTGGAAAGGCGAGCCCGAGCATCTTCCAGAGAGCATGACCAAGATCCTGG 1177  
961 Db TGCCCCAGGGCTGGAAAGGCGAGCCCGGCATCTTCCAGAGAGCATGACCAAGATCCTGG 1020  
1178 QY AGCCCTTCGGCGCCCGCAACCCCGAGATCGTGATCTTACAGGCCCGCCCTGTACGTTGGCA 1237  
1021 Db AGCCCTTCGGCAAGCAGAACCCCGACATCGTGATCTTACAGGCCCGCCCTGTACGTTGGCA 1080  
1238 QY GCGACCTGGAGATCGGCCAGCACCGCGCCAGATCGAGAGCTGCGCAAGCATCTGTCG 1297  
1081 Db GCGACCTGGAGATCGGCCAGCACCGCCACCAAGATCGAGAGCTGCGCCAGCATCTGTCG 1140  
1298 QY GCTGGGGCTTACACACCCCGACAGAGACCAAGAGAGCCCGCCCTTCTGTGGATGG 1357  
1141 Db GCTGGGGCTTACACACCCCGACAGAGACCAAGAGAGCCCGCCCTTCTGTGGATGG 1200  
1358 QY GCTACGAGCTGCACCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGGAGA 1417  
1201 Db GCTACGAGCTGCACCCGACAAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGGAGA 1260  
1418 QY GCTGGACCTGAAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGTCT 1477  
1261 Db GCTGGACCTGAAACGACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGTCT 1320  
1478 QY ACCCGGATCAAGTGTGCCAGCTGTGCAAGCTGCTCGGGGCGCAAGGCCCTGACCG 1537  
1321 Db ACCCGGATCAAGTGTGCCAGCTGTGCAAGCTGCTCGGGGCGCAAGGCCCTGACCG 1380  
1538 QY ACATCGTCCCTGACCCGAGGAGGCGGAGCTGGAGCTGCCGAGAACCGGAGATCTCTGC 1597  
1381 Db AGGTGATCCCTGACCCGAGGAGGCGGAGCTGGAGCTGCCGAGAACCGGAGATCTCTGA 1440  
1598 QY GCGAGCCGTGCAAGGCGGTGTACTAGACCCGAGCAAGGACTGTGTGCCAGATCCAGA 1657  
1441 Db AGGAGCCGTGCAAGGCGGTGTACTAGACCCGAGCAAGGACTGTGTGCCAGATCCAGA 1500  
1658 QY AGCAGGCGCAGCAGCAGTGGACCTACAGATCTACAGAGGCGCTTCAAGAACCTGAAGA 1717  
1501 Db AGCAGGCGCAGGCGCAGTGGACCTACAGATCTACAGAGGCGCTTCAAGAACCTGAAGA 1560  
1718 QY CCGCAAGTACGCCAAGATGGCGACCGCCACCAACAGAGTGAAGCAGCTGACCCGAGG 1777  
1561 Db CCGCAAGTACGCCCAGATGGCGCGCGCCACACCAACAGAGTGAAGCAGCTGACCCGAGG 1620  
1778 QY CCGTGCAGAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCC 1837  
1621 Db CCGTGCAGAGATCGACACCGAGAGCATCGTGATCTGGGGCAAGATCCCAAGTTCAAGC 1680  
1838 QY TGCCCATCCAGAGAGAGACTGGGAGACCTGGTGACCGACTACTTGGCAGGCGCCACCTGGA 1897  
1681 Db TGCCCATCCAGAGAGAGACTGGGAGGCGCTGGTGATGGAGTACTTGGCAGGCGCCACCTGGA 1740  
1898 QY TCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGA 1957  
1741 Db TCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGA 1800  
1958 QY AGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAACCGCGCCCGCCCAACCGCGAGACA 2017  
1801 Db AGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAACCGCGCGCCCGCCCAACCGCGAGACA 1860  
2018 QY AGATCGGCAAGCGCGCTACGTGACCGGACCGGCGCGCAGAGATCGTGAAGCTGACCG 2077  
1861 Db AGTGGGCAAGCGCGCTACGTGACCGGACCGGCGCGCAGAGGTTGTGAGCATCGCGC 1920  
2078 QY AGACCACCAACAGAGAGCCGAGCTGACGAGCCCATCGAGCTGCGCCCTGAGGACAGCGGCA 2137  
1921 Db ACACCACCAACAGAGAGCCGAGCTGACGAGCCCATCGAGCTGCGCCCTGAGGACAGCGGCA 1980  
2138 QY CCGAGGTGAACATCGTGACCGGACAGCAGTACGCGCTTGGGATCATTCAGAGGCCGAGCCG 2197

1981 Db TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCAGGCCCGAGCCG 2040  
2198 QY ACAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2257  
2041 Db ACAAGAGCGAGAGCGAGCTGGTGAAGCCAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2100  
2258 QY TGTACTAGCTGGGTGCGCCGCCCAAGGGGATCGCGGGCAACAGAGAGATCGACAAGC 2317  
2101 Db TGTACTGGCTGGGTGCGCCGCCCAAGGGGATCGCGGGCAACAGAGAGGTTGGAAGC 2160  
2318 QY TGGTGAGCAAGCGGATCGCAAGTGTCTTCTGAGCGGATCGATCGCGCATCTGGA 2377  
2161 Db TGGTGAGCGCGGATCGCAAGTGTCTTCTGAGCGGATCGATCGCGCATCTGGA 2220  
2378 QY TCTACAGTACATGAGCAGACCTGTACGTGGGCGAGCGCGGCTTAGGATCGATTAAAGC 2437  
2221 Db TCTACAGTACATGAGCAGACCTGTACGTGGGCGAGCGCGGCTTAGGATCGATTAAAGC 2280  
2438 QY TTCCCGGGCTAGCACCGGTGAATTC 2463  
2281 Db TTCCCGGGCTAGCACCGGTGAATTC 2306

RESULT 7  
AX427930 9166 bp DNA linear PAT 20-JUN-2002  
LOCUS Sequence 168 from Patent W00232943.  
ACCESSION AX427930  
VERSION AX427930.1 GI:21538017  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Huang, Y. and Nabel, G. J.  
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization  
JOURNAL Patent: WO 0232943-A 168 25-APR-2002;  
GOVERNMENT OF THE UNITED STATES (US)  
FEATURES Location/Qualifiers  
source 1..9166  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="plasmid pVR1012x/s containing HIV genes"

Query Match 82.3%; Score 2027.6; DB 6; Length 9166;  
Best Local Similarity 91.7%; Pred. No. 6.6e-203;  
Matches 2194; Conservative 0; Mismatches 179; Indels 19; Gaps 4;

QY 14 TGGCCGAGGCGCATGAGCGAGCCAGCCAGCGCCAAACATCTGATCGAGCGCAGCACTTCA 73  
Db 2961 TGGCCGAGGCGCATGAGCAAGGTGAACCAACATCATGATGAGCGCAGCACTTCA 3020  
QY 74 AGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCGCCACATCGCCGCA 133  
Db 3021 AGGCGCCCAAGCGCACCATCAAGTGTCTTCAACTGCGGCAAGGAGGCGCCACCTGGCCGCA 3080  
QY 134 ACTCGCGCGCCCGCGCAAGAGGGCTGTCTGGAAGTGGGCAAGGAGGCGCCACAGATGA 193  
Db 3081 ACTCGCGCGCCCGCGCAAGAGGGCTGTCTGGAAGTGGGCAAGGAGGCGCCACAGATGA 3140  
QY 194 AGGACTGACCGCGCGCGCGCGCAAGTCTTCTCGGAGGAGCTGCGCTTCCCGCGAGGCA 253  
Db 3141 AGGACTGACCGCGCGCGCGCGCAAGTCTTCTCGGAGGAGCTGCGCTTCCCGCGAGGCA 3199  
QY 254 AGGCGCGCGAGTCTCCCGAGCGAGCAAGCCCGCGCAACAGCGCCCAACAGCGCGCGAGTGC 313  
Db 3200 AGGCGCGAGATTTTCTTCAGGCGAGCAGCAGCCACAGCCCGCCACAGAGAGAGCTTC 3259  
QY 314 AGGTGCGCGG-----CGACAAACCCCGCAGGAGCGCGCGCGCGCGCGCGCGCA--- 364

Db 3260 AGGTTTGGGAAGAGACAACAACCTCCCTCTCAGAAAGCAGGAGCCGATAGACAAGAACTG 3319  
 Qy 365 ---CCCTGAACCTTCCCGCAGATCAACCTGTGGAGAGCGCCCTCGTGAGCATCAAGGTGG 421  
 Db 3320 TATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCAGCCCTCGTCAACAATAAGATAG 3379  
 Qy 422 GCAGCGAGATCAAGAGGCGCTGTGGACACCGGCGGCGGAGACACCGTGTGGAGGAGA 481  
 Db 3380 GGGGCCAGCTGAAGAGGCGCTTTAGACACCGGCGCGGAGCAGACCGCTGTGGAGAGA 3439  
 Qy 482 TGAGCCTGCCCGGCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCAATCAAGG 541  
 Db 3440 TGAACCTGCCCGGCGCTGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCAATCAAGG 3499  
 Qy 542 TGCGCCAGTACGACCAAGATCTGTATCGAGATCTCGGCAAGAGCCGATCGGCAACCGTGC 601  
 Db 3500 TGGGCCAGTACGACCAAGATCTGTATCGAGATCTCGGCAAGAGCCGATCGGCAACCGTGC 3559  
 Qy 602 TGATCGGCGCCACCGCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCTGCA 661  
 Db 3560 TGGTGGGCGCCACCGCGTGAACATCATCGGCGCAACATGCTGACCCAGATCGGCTGCA 3619  
 Qy 662 CCCTGAACCTTCCCATCAGCCCCCATCGAGACCGGTGCGGCAAGCTGAAGCCCGGCTG 721  
 Db 3620 CCCTGAACCTTCCCATCAGCCCCCATCGAGACCGGTGCGGCAAGCTGAAGCCCGGCTG 3679  
 Qy 722 ACGCCCCCAAGGTGAAGCAGTGGCCCTTGACCGAGAGAGATCAAGGCCCTGACCCGCA 781  
 Db 3680 ACGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGAGAGATCAAGGCCCTGAGTGGAGA 3739  
 Qy 782 TCTCGAGGAGATGGAGAGAGGCGGCAAGATCAACAGATCGGCGGCAAGCTGGTGGACT 841  
 Db 3740 TCTGCAACGAGATGGAGAGAGGCGGCAAGATCAACAGATCGGCGGCAAGCTGGTGGACT 3799  
 Qy 842 ACACCCCGTGTTCGCCATCAAGAGAGAGGCAAGATCAACAGATCGGCGGCAAGCTGGTGGACT 901  
 Db 3800 ACACCCCGTGTTCGCCATCAAGAGAGAGGCAAGATCAACAGATCGGCGGCAAGCTGGTGGACT 3859  
 Qy 902 TCCGCGAGCTGAAACAAGCGCAACCAAGCTTCTGGAGGTGACGTGGGCAATCCCCAACC 961  
 Db 3860 TCCGCGAGCTGAAACAAGCGCAACCAAGCTTCTGGAGGTGACGTGGGCAATCCCCAACC 3919  
 Qy 962 CCGCGGCGCTGAAGAGAGAGAGCTGACCGGTCTGGAGGTGAGCGGCGGAGCCCTACTTCA 1021  
 Db 3920 CCGCGGCGCTGAAGAGAGAGAGCTGACCGGTCTGGAGGTGAGCGGCGGAGCCCTACTTCA 3979  
 Qy 1022 GCCTGCCCTTGACGAGGACTTCCGCAAGTACACCGCTTCAACCATCCCCAGCATCAACA 1081  
 Db 3980 GCCTGCCCTTGACGAGGACTTCCGCAAGTACACCGCTTCAACCATCCCCAGCATCAACA 4039  
 Qy 1082 ACAGAGACCCCGGATCCGCTACAGTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGCGAGCC 1141  
 Db 4040 ACAGAGACCCCGGATCCGCTACAGTACAAAGTGTGCCCCAGGGCTGGAAGGCGAGCC 4099  
 Qy 1142 CCAGCATCTTCCAGAGCAGATGACCAAGATCTCGAGGCGCTTCGCGCGCCGCAACCCCG 1201  
 Db 4100 CCGCATCTTCCAGTGAAGCATGACCAAGATCTCGAGGCGCTTCGCAAGCAGAACCCCG 4159  
 Qy 1202 AGATCGTGATCTACCA-----GGCCCCCTGTGACGTGGGCGAGGACCTGGAGATCGGCC 1255  
 Db 4160 ACATCGTGATCTACCAAGTACATGAGCCACCTGTGACGTGGGCGAGGACCTGGAGATCGGCC 4219  
 Qy 1256 AGCAACCGGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGCTTTCACACCC 1315  
 Db 4220 AGCAACCGCACAAGATCGAGAGCTGCGCCAGCACCTGCTGCGCTGGGCTTTCACACCC 4279  
 Qy 1316 CCGCACAAGCAGCAGAGAGAGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCG 1375  
 Db 4280 CCGCACAAGAGCAGCAGAGAGAGCCCTTCTCTGTGATGGGCTACGAGCTGCACCCCG 4339  
 Qy 1376 ACAAGTGACCGTGCAGGCCCATCGAGCTGCCCCGAGAGAGAGCTGGACCGTGAACGACA 1435

Db 4340 ACAAGTGGACGCTGCAGCCCATCGTGTGCTGCCGAGAGAGGACGCTGTGACCGTGAACGACA 4399  
 Qy 1436 TCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCGGGATCAAGGTGC 1495  
 Db 4400 TCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCGGGATCAAGGTGC 4459  
 Qy 1496 GCAGAGCTGTCAAGCTGCTGCGGGCGCCCAAGGCCCTTGACCGACATCGTGCCTTGACCG 1555  
 Db 4460 GCAGAGCTGTCAAGCTGCTGCGGGCACCAAGGCCCTTGACCGAGGTGTGCTGCGCTGACCG 4519  
 Qy 1556 AGAGGCGGAGCTGGAGTGGCGGAGAAACCGGAGATCTTGGCGGAGCCGCTGTCACGGCG 1615  
 Db 4520 AGAGGCGGAGCTGGAGTGGCGGAGAAACCGGAGATCTTGAAGGAGCCCGCTGTCACGGCG 4579  
 Qy 1616 TGTACTACGACCCCGCAGCAAGGACCTGTGTGGCGGAGATCCAGAGCAGGCGCACACCAAGT 1675  
 Db 4580 TGTACTACGACCCCGCAGCAAGGACCTGTATCGCGGAGATCCAGAGCAGGCGCAGGGCCAGT 4639  
 Qy 1676 GGACTTACCAGATCTTACGAGGAGCCCTTCAAGAACCTTGAAGACCGCGCAAGTACGCCAAGA 1735  
 Db 4640 GGACTTACCAGATCTTACGAGGAGCCCTTCAAGAACCTTGAAGACCGCGCAAGTACGCCGCA 4699  
 Qy 1736 TGGCAGCCGCGCCACACAAAGAGAGTGAAGCAGCTGACCGAGGCGCTGCGAGAGATCGCCA 1795  
 Db 4700 TGAAGGCGCGCCACACAAAGAGAGTGAAGCAGCTGACCGAGGCGCTGCGAGAGATCGCCA 4759  
 Qy 1796 TGGAGAGCATGTGATCTGGGCGAAGACCCCAAGTTCCGCTGCCCATCCAGAGAGAGA 1855  
 Db 4760 CCGAGAGCATGTGATCTGGGCGAAGACCCCAAGTTCAAGCTGCCCATCCAGAGAGAGA 4819  
 Qy 1856 CTTGGGAGAGCATGTGTGACCGGACTTACTGTGCGAGGCGCACTTGATCCCGAGTGGGAGTTGC 1915  
 Db 4820 CTTGGGAGGCTGTGTGACCGGACTTACTGTGCGAGGCGCACTTGATCCCGAGTGGGAGTTGC 4879  
 Qy 1916 TGAACACCCCGCTGTGTGAAAGCTGTGTTGAACAGCTGTGAGAGAGAGCCCATCATCGCGC 1975  
 Db 4880 TGAACACCCCGCTGTGTGAAAGCTGTGTTGAACAGCTGTGAGAGAGAGCCCATCATCGCGC 4939  
 Qy 1976 CCGAGACTTCTACGTGAGCGGCGCGCAACCGCGAGACCAAGATCGGCAAGGCCCGCT 2035  
 Db 4940 CCGAGACTTCTACGTGAGCGGCGCGCAACCGCGAGACCAAGCTGGGCAAGGCCCGCT 4999  
 Qy 2036 ACCTGACCGGAGCGGCGCGCAGAGATCGTGTGAGCCTTGAGCCGAGACCAACCAAGAGAGA 2095  
 Db 5000 ACCTGACCGGAGCGGCGCGCAGAGAGTGTGTGCGCTTGAACCGACACCAACCAAGAGAGA 5059  
 Qy 2096 CCGAGCTGCGAGGCGCATCCAGCTGCGCTTGCAGGACAGCGGCGAGCGGTGAACATCGTGA 2155  
 Db 5060 CCGAGCTGCGAGGCGCATCCAGCTGCGCTTGCAGGACAGCGGCGCTGAGGTGAACATCGTGA 5119  
 Qy 2156 CCGACAGCGAGTACCGCTGGGCGATCATCCAGGCGCCAGCCCGCAAGAGCGGAGAGCGAGC 2215  
 Db 5120 CCGACAGCGAGTACCGCTGGGCGATCATCCAGGCGCCAGCCCGCAAGAGCGGAGAGCGAGC 5179  
 Qy 2216 TGTGTAAACCAAGATCATCGAGCAGCTGTATCAAGAGAGAGAGGTGTACTGTGAGCTGGGTGC 2275  
 Db 5180 TGTGTAGCCAGATCATCGAGCAGCTGTATCAAGAGAGAGAGGTGTACTGTGCGCTGGGTGC 5239  
 Qy 2276 CCGCCCAACAAGGCGATCGGCGGCAACCGAGCAGATCGACAAGCTGTGTGAGCAGGCGCATCC 2335  
 Db 5240 CCGCCCAACAAGGCGATCGGCGGCAACCGAGCAGGTGTGAGCGGCTGTGAGCGCCCGCATCC 5299  
 Qy 2336 GCAAGGTGCTGTTCTCGAGCGGCGATCGATGGGCGCATCGTGATCTACCAAGTGA 2387  
 Db 5300 GCAAGGTGCTGTTCTCGAGCGGCGATCGACAAGGCCCGAGGAGGAGCAGCAAGNA 5351

RESULT 8  
 BD263706  
 LOCUS BD263706 2312 bp DNA linear PAT 17-JUL-2003  
 DEFINITION Improved expression of HIV polypeptides and production of virus-like particles.  
 ACCESSION BD263706



```

VERSION      BD263706.1 GI:33073474
KEYWORDS     JP 2002533124-A/73.
SOURCE       synthetic construct
ORGANISM     other sequences; artificial sequences.
REFERENCE    1 (bases 1 to 2312)
AUTHORS      Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
              Greer,C., Selby,M. and Walker,C.
TITLE        Improved expression of HIV polypeptides and production of
              virus-like particles
JOURNAL      Patent: JP 2002533124-A 73 08-OCT-2002;
              CHIRON CORP
COMMENT      OS Artificial Sequence
              PN JP 2002533124-A/73
              PD 08-OCT-2002
              PF 30-DEC-1999 JP 2000591193
              PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
              SUSAN BARNETT,JAN ZUR MEDEDE,INDRESH SRIVASTAVA,YING LIAN, PI
              KARIN HARTOG,
              PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
              C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
              C12N5/10,
              PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
              Description of Artificial Sequence: FS(-).protmod.RTopt (+) FH Key
              Location/Qualifiers
FT source    1. .2312
              /organism='Artificial Sequence'.
FT           Location/Qualifiers
              1. .2312
              /organism='synthetic construct'
              /mol_type='genomic DNA'
              /db_xref='taxon:32630'

ORIGIN
Query Match      82.2%; Score 2025.2; DB 6; Length 2312;
Best Local Similarity 93.3%; Pred.No.1.7e-202;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGCAAGAGGGGCCACACGATGAAGACTGCACCGAGCGCCAGGCAACTTCTTCCGCG 229
DB 1 GCGGCGCGGAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60

QY 230 AGGACCTGGCTTCCCGAGGCAAGGCCCGCGAGTTCGCCAGCAGAGCAACCGGCA 289
DB 61 AGGACCTGGCTTCTTGCAGGCGCAAGGCCCGCGAGTTACAGCAGCAGCAGCAGCCGCGCA 120

QY 290 ACAGCCCCACACCGCGAGCTGCAGGTGCGCGCG- - - - -ACAAACCCCGCAGCGAGG 343
DB 121 ACGCCCCACACCCCGCGAGCTGCAGGTGCGCGCGGCGAGAACACAGCCTGAGCGAGG 180

QY 344 CCGGCGCCGAGCGCCAGGCGCACCTTG- - - - -AACTTCCCCCAGATCACCTTGTGGCAGC 397
DB 181 CCGGCGCCGAGCGCCAGGCGCACCGTGAGCTTCAACTTCCCCCAGATCACCTTGTGGCAGC 240

QY 398 GCCCCTTGTGAGCATCAAGTGTGGCGGCGCAGATCAAGAGGCCCTGTGTGACACCGGCG 457
DB 241 GCCCCTTGTGACCATCAAGTGTGGCGGCGCAGCTCAAGAGGCGCTGTGTGACACCGGCG 300

QY 458 CCGAGCACACCGTGTGAGGAGATGAGCTTCCCGCAAGTGAAGCCCAAGATGATCG 517
DB 301 CCGAGCACACCGTGTGAGGAGATGAACTTCCCGCAAGTGAAGCCCAAGATGATCG 360

QY 518 GCGGATCGGCGGCTTCAATCAAGTGTGCGCAGTACGACCAAGATCCTGTGATCGAGATCTGCG 577
DB 361 GCGGATCGGCGGCTTCAATCAAGTGTGCGCAGTACGACCAAGATCCTGTGATCGAGATCTGCG 420

QY 578 GCAAGAGCCCATCGGCGCCGCTGTGATCGGCGCCCAACCCCGTGAACATCATTCGCGCGCA 637
DB 421 GCCAAGAGCCCATCGGCGCCGCTGTGATCGGCGCCCAACCCCGTGAACATCATTCGCGCGCA 480

QY 638 ACATGCTACCCAGCTGGGCTGCACCTGAACTTCCCATCAGCCCATCGAGCCGCTGC 697
DB 481 ACCTGCTACCCAGATCGGCTGCACCTGAACTTCCCATCAGCCCATCGAGCCGCTGC 540

698 CCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTACCGAGG 757
541 CCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTACCGAGG 600
758 AGAAGATCAAGGCCCTGAGCCGCATCTGCGAGGAGTGAAGAGGAGGCGCAGATCACCA 817
601 AGAAGATCAAGGCCCTGAGCCGCATCTGAGCGAGTGAAGAGGAGGCGCAGATCAGCA 660
818 AGATCGGCCCCGAGAACCCCTACAACACCCCGGTGTTCGCCATCAAGAAAGAGGACAGCA 877
661 AGATCGGCCCCGAGAACCCCTACAACACCCCGGTGTTCGCCATCAAGAAAGAGGACAGCA 720
878 CCAAGTGGCGCAAGCTGTGTGAGCTTCCGCGAGCTGAAACAAGCGCACCCAGAGATCTTGGG 937
721 CCAAGTGGCGCAAGCTGTGTGAGCTTCCGCGAGCTGAAACAAGCGCACCCAGAGATCTTGGG 780
938 AGGTGCAGTGGGCATCCCGCACCCCGCGCTGAAGAGAGAGAGAGCGGTGACCGTGC 997
781 AGGTGCAGTGGGCATCCCGCACCCCGCGCTGAAGAGAGAGAGAGCGGTGACCGTGC 840
998 TGGAGTGGCGCAGCGCTACTTTCAGCGTGGCCCTGAGCAGGACTTTCGCCAAGATCACCG 1057
841 TGGAGTGGCGCAGCGCTACTTTCAGCGTGGCCCTGAGCAGGACTTTCGCCAAGATCACCG 900
1058 CTTTCAACATCCCGAGCATCAACAACAGAGACCCCGCGCATCCGCTACCAAGTCAACGTC 1117
901 CTTTCAACATCCCGAGCATCAACAACAGAGACCCCGCGCATCCGCTACCAAGTCAACGTC 960
1118 TGCCCCGGGCTGGAAGGGCAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTTGG 1177
961 TGCCCCGGGCTGGAAGGGCAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCTTGG 1020
1178 AGCCCTTCCGCGCCGCAACCCCGAGATCGTATCTACCA- - - - -GGCCCCCTGTGACG 1231
1021 AGCCCTTCCGAGCAGAACCCCGAGCATCTGATCTTCCAGTACATGAGCAGCTGTGACG 1080
1232 TGGCAGCGCACTTGAGATCGGCCAGCACCGCGCAAGATCGAGAGTGTGCGCAAGCACC 1291
1081 TGGCAGCGCACTTGAGATCGGCCAGCACCGCGCAAGATCGAGAGTGTGCGCGAGCACC 1140
1292 TGCTGCGTGGGCTTACACACCCCGAGCAGTGTGAGTGTGAGCGCTGAGCGTGTGCGGAG 1351
1141 TGCTGCGTGGGCTTACACACCCCGAGCAGTGTGAGTGTGAGCGCTGAGCGTGTGCGGAG 1200
1352 GGATGGGCTTACGAGTGTGACACCCCGAGTGTGAGTGTGAGCGCTGAGCGTGTGCGGAG 1411
1201 GGATGGGCTTACGAGTGTGACACCCCGAGTGTGAGTGTGAGCGCTGAGCGTGTGCGGAG 1260
1412 AGGAGAGTGAACCGTGAACGACATCCAGAAAGCTGTGGGCAAGTGAATCGGCGCAGCC 1471
1261 AGGAGAGTGAACCGTGAACGACATCCAGAAAGCTGTGGGCAAGTGAATCGGCGCAGCC 1320
1472 AGATCTACCCCGCATCAAGTGTGCGCAGCTGTGCAAGTGTGCGCGCGCCAGAGGCC 1531
1321 AGATCTACCCCGCATCAAGTGTGAGTGTGCAAGTGTGCAAGTGTGCGCGCGCCAGAGGCC 1380
1532 TGACCGACATCGTGTGCGCCGAGCAGGAGCGCGAGCTGAGCTGTGCGCGAGAACCGGAG 1591
1381 TGACCGAGTGTATCCCTCTGACCGAGGAGCGAGCTGAGCTGTGCGCGAGAACCGGAG 1440
1592 TCCTGCGGAGCCCGTGTGACCGCGTGTACTACGACCCAGCAAGGACCTTGTGGCGGAG 1651
1441 TCCTGAAGGAGCCCGTGTGACCGAGGTGTACTACGACCCAGCAAGGACCTTGTGGCGGAG 1500
1652 TCCAGAACAGGCGCACGACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1711
1501 TCCAGAACAGGCGCGAGGCGCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1560
1712 TGAAGACCGGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1771
1561 TGAAGACCGGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1620
```



1772 CCAGGCGCTGCAGAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGT 1831  
Db |||||  
1621 CCGAGGCGGTGCAGAGGTGAGCAACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT 1680  
Qy |||||  
1832 TCCGCTCTGCCCATCCAGAAGGAGACCTTGGGAGACCTTGGTGAGACCGACTACTTGGCAGGCCA 1891  
Db |||||  
1681 TCAAGCTTGCCCATCCAGAAGGAGACCTTGGGAGGCTTGGTGAGTGAGTACTTGGCAGGCCA 1740  
Qy |||||  
1892 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTTGTGTAAGCTGTGGTACAGC 1951  
Db |||||  
1741 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTTGTGTAAGCTGTGGTACAGC 1800  
Qy |||||  
1952 TGGAGAAGGAGCCCATCATCGGCGCGAGACTTCTAGCTGAGACCGCGCGCCCAACCGCG 2011  
Db |||||  
1801 TGGAGAAGGAGCCCATCTGTGGCGCGAGACTTCTAGCTGAGACCGCGCGCGCCCAACCGCG 1860  
Qy |||||  
2012 AGACCAAGATCGGCAAGCGCGCTTACGTGACCGACAGCCAGTACCGCGCGCGCGCCCAACCGCG 2071  
Db |||||  
1661 AGACCAAGCTGGGCAAGCGCGCTTACGTGACCGACAGCCAGTACCGCGCGCGCGCCCAACCGCG 1920  
Qy |||||  
2072 TGACCGAGACACCAACAGAGACCGAGCTGAGGAGCATCCAGCTGGCCCTGCAGGACA 2131  
Db |||||  
1921 TCGCCGACACCAACAGAGACCGAGCTGAGGAGCATCCAGCTGGCCCTGCAGGACA 1980  
Qy |||||  
2132 GCGGACGAGGTGAACATCGTGACCGACAGCCAGTACCGCTGGGATCATCCAGGCC 2191  
Db |||||  
1981 GCGGCTTGGAGTGAAACATCGTGACCGACAGCCAGTACCGCTGGGATCATCCAGGCC 2040  
Qy |||||  
2192 AGCCCGCAAGAGCGAGCGAGCTGGTGAAACAGATCATCGAGCAGCTGATCAAGAAGG 2251  
Db |||||  
2041 AGCCCGACAGAGCGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100  
Qy |||||  
2252 AGAAGTGTACCTGAGCTGGTGCCCGCCCAAGGAGCATCGGCGGCAACAGAGATCG 2311  
Db |||||  
2101 AGAAGTGTACCTGGCTTGGTGCCCGCCCAAGGAGCATCGGCGGCAACAGAGAGTGG 2160  
Qy |||||  
2312 ACAAGCTGTGAGCAAGGAGCATCGCAAGGTGCTGTTCTTGGAGGAGCATCGATGGCGGCA 2371  
Db |||||  
2161 ACAAGCTGTGAGCGCGGCATCCGCAAGGTGCTGTTCTTGGAGGAGCATCGATGGCGGCA 2220  
Qy |||||  
2372 TCGTGATCTACAGTACATGAGACGACCTGTACGTGGGAGCGCGGCTTAGGATCGATT 2431  
Db |||||  
2221 TCGTGATCTACAGTACATGAGACGACCTGTACGTGGGAGCGCGGCTTAGGATCGATT 2280  
Qy |||||  
2432 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2463  
Db |||||  
2281 AAAAGCTTCCGGGCTAGCACCGGTGAATTC 2312

RESULT 9  
CQ870576 2312 bp DNA linear PAT 13-SEP-2004  
LOCUS  
DEFINITION Sequence 84 from Patent EP143851.  
ACCESSION CQ870576  
VERSION CQ870576.1 GI:52000092  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM other sequences; artificial sequences.  
REFERENCE 1  
AUTHORS Barnett,S., Zurmege,J., Srivastava,I., Lian,Y., Hartog,K.,  
Liu,H., Greer,C., Selby,M. and Walker,C.  
TITLE Improved expression of HIV polypeptides and production of  
virus-like particles  
JOURNAL Patent: EP 143851-A 84 30-JUN-2004;  
CHIRON CORPORATION (US)  
FEATURES  
source  
1..2312  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Description of Artificial Sequence:  
FS (-). protmod.RTopt (+)"

ORIGIN  
Query Match 82.2%; Score 2025.2; DB 6; Length 2312;  
Best Local Similarity 93.3%; Pred. No. 1.7e-202;  
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
Qy 170 GCGGCAAGGAGGCGCCACAGATGAAGACTGCAACCGAGCGCCAGGCCAACTTCTTCCGCG 229  
Db |||||  
1 GCGGCGCGAAGGACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60  
Qy 230 AGAAGCTTGGCTTCCCGCAGGCAAGGCGCGAGTTTCCCGAGGAGCAACCGCGCCA 289  
Db |||||  
61 AGAAGCTTGGCTTCTTCAAGGCAAGGCGCGAGTTTCAAGGAGGAGGCTTCTTCCGCGCCA 120  
Qy 290 ACAGCCCCACAGCGGAGCTGCAGGTGCGCGCGG-----ACAACCCCGCAGCGAGG 343  
Db |||||  
121 ACAGCCCCACCGCGCGAGCTGCAGGTGTGGGCGCGGAGAACACAGCTTGAGCGAGG 180  
Qy 344 CCGGCGCGAGCGCCAGGCGCACCTG-----AACTTCCCGCCAGATCACCTTGTGGCAGC 397  
Db |||||  
181 CCGGCGCGAGCGCCAGGCGCACCTGAGCTTCAACTTCCCGCCAGATCACCTTGTGGCAGC 240  
Qy 398 GCGGCTTGTGAGATCAAGGTGGCGCGCAGATCAAGGAGGCGCTTCTGTGAACCGCGCG 457  
Db |||||  
241 GCGGCTTGTGAGATCAAGGTGGCGCGCAGCTCAAGGAGGCGCTTCTGTGAACCGCGCG 300  
Qy 458 CCGACGACCGGTGCTGGAGGAGTGAAGCTGCGCGCGCAAGTGGAGGCCCAAGATGATCG 517  
Db |||||  
301 CCGACGACCGGTGCTGGAGGAGTGAAGCTGCGCGCGCAAGTGGAGGCCCAAGATGATCG 360  
Qy 518 GCGGATCGGCGGCTTCAAGGTGGCGCCAGTACGACAGATCCTGATCGAGATCTTGGCG 577  
Db |||||  
361 GCGGATCGGCGGCTTCAAGGTGGCGCAGTACGACAGATCCTGAGATCTTGGCG 420  
Qy 578 GCAAGAGGCGCATCGGCAACCGTGTGATCGGCGCGCCCAAGGTGAAGCAGTGGCGCGCGCA 637  
Db |||||  
421 GCAAGAGGCGCATCGGCAACCGTGTGATCGGCGCGCGCCCAAGGTGAAGCAGTGGCGCGCA 480  
Qy 638 ACATGTGACCGAGCTGGGCTGCACCTGAATCTCCCATCAGCCCATCGAGACCGTGC 697  
Db |||||  
481 ACCTGTGACCGAGATCGGCTGCACCTGAATCTCCCATCAGCCCATCGAGACCGTGC 540  
Qy 698 CCGTGAAGCTGAAGCGCGCATGAGCGCGCCCAAGGTGAAGCAGTGGCGCGCGCGCGCG 757  
Db |||||  
541 CCGTGAAGCTGAAGCGCGCATGAGCGCGCCCAAGGTGAAGCAGTGGCGCGCGCGCGCG 600  
Qy 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGAGAGGAGGCGCGAGATCACCA 817  
Db |||||  
601 AGAAGATCAAGGCGCTGATGAGATCTGACCGGAGATGAGAGGAGGCGCGAGATCACCA 660  
Qy 818 AGATCGGCGCGGAGACCCCTTACACACCCCGCTGTTTCGCCATCAAGAAGAGGACAGCA 877  
Db |||||  
661 AGATCGGCGCGGAGACCCCTTACACACCCCGCTGTTTCGCCATCAAGAAGAGGACAGCA 720  
Qy 878 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937  
Db |||||  
721 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780  
Qy 938 AGGTGAGCTGGGCGATCCCGCACCCCGCGGCTGAAAGAGAGAGAGCGGTGACCGTGC 997  
Db |||||  
781 AGGTGAGCTGGGCGATCCCGCACCCCGCGGCTGAAAGAGAGAGAGCGGTGACCGTGC 840  
Qy 998 TGGAGCTGGGCGAGCGCTTCTGAGCTGCGCGCGGAGGACTTCCGCAAGTACACCG 1057  
Db |||||  
841 TGGAGCTGGGCGAGCGCTTCTGAGCTGCGCGCGGAGGACTTCCGCAAGTACACCG 900  
Qy 1058 CTTTACCATCCCGAGCATCAACAACGAGACCCCGCGCATCCCGTACAGTACACCGTGC 1117  
Db |||||  
901 CTTTACCATCCCGAGCATCAACAACGAGACCCCGCGCATCCCGTACAGTACACCGTGC 960  
Qy 1118 TGCCCCAGGCTGAAGGCGAGCGCCCGAGCATCTTTCAGAGGAGCATGACCAAGATCTTGG 1177  
Db |||||  
961 TGCCCCAGGCTGAAGGCGAGCGCCCGCGCATCTTTCAGAGGAGCATGACCAAGATCTTGG 1020

```
Qy 1178 AGCCCTTCGCGCCGCGCAACCCCGAGATCGTGATCTACCA-----GCCCCCTGTACG 1231
Db 1021 AGCCCTTCGCAAGCAGAACCCCGACATCGTGTCTACAGTACATGACACCTGTACG 1080
Qy 1232 TGGGACGACCTGGAGATCGGCCAGCACCGCCCAAGATCAGGAGCTGCGCAAGCAC 1291
Db 1081 TGGGACGACCTGGAGATCGGCCAGCACCGCCCAAGATCAGGAGCTGCGCAAGCAC 1140
Qy 1292 TGCTGCGCTGGGGCTTCCACCCCGCAAGAAGCACAGAGAGCCCCCTTCCTGT 1351
Db 1141 TGCTGCGCTGGGGCTTCCACCCCGCAAGAAGCACAGAGAGCCCCCTTCCTGT 1200
Qy 1352 GGATGGGCTACGAGCTGACCCCGCAAGTGAACCGTGACGCCATCAGCTGCCCGAGA 1411
Db 1201 GGATGGGCTACGAGCTGACCCCGCAAGTGAACCGTGACGCCATCAGCTGCCCGAGA 1260
Qy 1412 AGGAGAGCTGACCGTGAACACATCCAGAGAGCTGTGGGCAAGCTGAACCTGGCCAGCC 1471
Db 1261 AGGAGAGCTGACCGTGAACACATCCAGAGAGCTGTGGGCAAGCTGAACCTGGCCAGCC 1320
Qy 1472 AGATCTACCCCGGATCAAGGTGGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCC 1531
Db 1321 AGATCTAGCCCGGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCC 1380
Qy 1532 TGACCGACATCGTGCCTCGACCGAGGAGGCGAGCTGGAGCTGGCCGAGAACCCGAGGA 1591
Db 1381 TGACCGAGTGTATCCCTCGACCGAGGAGGCGAGCTGGAGCTGGCCGAGAACCCGAGGA 1440
Qy 1592 TCCTGCGGAGCGCTGACCGGCTGTACTAGACCCCGCAAGCAGCTGTGGTGGCCGAGA 1651
Db 1441 TCCTGAAGGAGCGCTGACCGGCTGTACTAGACCCCGCAAGCAGCTGTGGTGGCCGAGA 1500
Qy 1652 TCCAGAGCAGGCGCACGACAGTGGACTTACAGATCTACAGAGAGCCCTTCAAGAAC 1711
Db 1501 TCCAGAGCAGGCGCACGAGCTGGACCTTACAGATCTACAGAGAGCCCTTCAAGAAC 1560
Qy 1712 TGAAGACGGCAAGTACGCCAAGATGCGCACCGGCCACACCAAGAGCTGAAGCAGCTGA 1771
Db 1561 TGAAGACGGCAAGTACGCCCGCATGCGCGCGGCCACACCAAGAGCTGAAGCAGCTGA 1620
Qy 1772 CCGAGGCGCTGACAGATCCCATGGAGAGCATGTGTATCTGGGGCAAGACCCCAAGT 1831
Db 1621 CCGAGGCGCTGACAGAGTGAACCCGAGAGCATGTGTATCTGGGGCAAGATCCCAAGT 1680
Qy 1832 TCCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGAACCGACTACTTGGCAGGCCA 1891
Db 1681 TCAAGCTGCCCATCCAGAGGAGACCTGGGAGGCTGTGTGATGGATCTTGGCAGGCCA 1740
Qy 1892 CCTGGATCCCGAGTGGAGTTCTGGAACACCCCGCCCTCGTGAAGCTGTGTACCAAGC 1951
Db 1741 CCTGGATCCCGAGTGGAGTTCTGGAACACCCCGCCCTCGTGAAGCTGTGTACCAAGC 1800
Qy 1952 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTAGTGAAGCGGCGCCGCAACCGCG 2011
Db 1801 TGGAGAGGAGCCCATCATGTGGCGCGAGACCTTCTAGTGAAGCGGCGCCGCAACCGCG 1860
Qy 2012 AGACCAAGATCGGCAAGCGCGCTAGCTGACCGACCGGCGCGGCGAGAGATCTGTAGCC 2071
Db 1861 AGACCAAGCTGGGCAAGCGCGCTAGCTGACCGACCGGCGCGGCGAGAGTGTGTAGCA 1920
Qy 2072 TGACCCGAGACCAACCCAGAGACCGAGCTGCGAGGCCATCCAGCTGGCCCTGAGGACA 2131
Db 1921 TCGCCGACACCAACCCAGAGACCGAGCTGCGAGGCCATCCAGCTGGCCCTGAGGACA 1980
Qy 2132 GCGGAGGAGGTGAACATCTGTGACCGACAGCAGTACGCCCTGGGCATCATCAGGCC 2191
Db 1981 GCGGCTGGAGGTGAACATCTGTGACCGACAGCAGTACGCCCTGGGCATCATCAGGCC 2040
Qy 2192 AGCCCGACAGCAGGAGCGAGCTGTGTGAACACAGATCATCAGCAGCTGATCAGAGG 2251
Db 2041 AGCCCGACAGCAGGAGCGAGCTGTGTGAACACAGATCATCAGCAGCTGATCAGAGG 2100
```

```
Qy 2252 AGAAGGTGTACTGAGCTGGTGCCTCCGCCCAAGAGGATCGCGGGCAACGAGCAGATCG 2311
Db 2101 AGAAGGTGTACTGAGCTGGTGCCTCCGCCCAAGAGGATCGCGGGCAACGAGCAGATCG 2160
Qy 2312 ACAAGCTGGTGAAGCAAGGGCATCCGCAAGTGTCTTCTCGACGGCATCGATGCGGCA 2371
Db 2161 ACAAGCTGGTGAAGCGCGGCATCCGCAAGTGTCTTCTGAAACGGCATCGATGCGGCA 2220
Qy 2372 TCGTGATCTACCACTATATGGAACACCTGTACGTGGGAGCGCGGCTTAGGATCGATT 2431
Db 2221 TCGTGTACTACCACTATATGGAACACCTGTACGTGGGAGCGCGGCTTAGGATCGATT 2280
Qy 2432 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2463
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 10
AR373389
LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett, S. W., Megede, J., Greer, C. and Selby, M.
TITLE Expression of HIV polypeptides and production of virus-like
particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
FEATURES
source
1. .2312
/organism="unknown"
/moi_type="genomic DNA"

Query Match 82.2%; Score 2025.2; DB 6; Length 2312;
Best Local Similarity 93.3%; Pred. No. 1.7e-202;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy 170 GCGCAAGAGGGGCCACAGATGAAGACTGCACCGAGCGCCAGCCAACTTCTTCCGCG 229
Db 1 GCGGCGCGAAGGACACCAAAATGAAGATTGCACTGAGAGACAGGCTAAATTTCTTCCGCG 60
Qy 230 AGGACCTGGCTTCCCGCAGGCAAGCCCGCAGTTCGCCAGGAGCAGAACCGCGCA 289
Db 61 AGGACCTGGCTTCTCGAGGCAAGCCCGCAGTTCAGCAGCGAGCAGAACCGCGCA 120
Qy 290 ACAGCCCCACAGCCGCGAGCTGCAGGTGCGCGCGC-----ACAACCCCCGAGCGAGG 343
Db 121 ACAGCCCCACCGCGCGAGCTGCAGGTGCGCGCGCGAGAAACAACAGCTGAGCGAGG 180
Qy 344 CCGGCGCGAGCGCGAGGCAACCCCTG-----AACTTCCCCCAGATCAACCTGTGGCAGC 397
Db 181 CCGGCGCGAGCGCGAGGCAACCCCTGAGCTTCAACTTCCCCCAGATCAACCTGTGGCAGC 240
Qy 398 GCGCCCTGTGTAGCATCAAGTGGCGGCCAGATCAAGGAGGCCCTGCTGCACACCGCGC 457
Db 241 GCGCCCTGTGTAGCATCAAGTGGCGGCCAGCTCAAGGAGGCCCTGCTGCACACCGCGC 300
Qy 458 CCGACGACACCGTCTCGAGGAGATGAGCTTCCCGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTCTCGAGGAGATGAACCTGCGCGGCAAGTGAAGCCCAAGATGATCG 360
Qy 518 GCGGATCGGCGGCTTCAATCAAGTGGCCAGTACGACACAGATCCTGATCAGATCTGCG 577
Db 361 GCGGATCGGCGGCTTCAATCAAGTGGCGGAGTACGACACAGATCCCGCTGGAGATCTGCG 420
Qy 578 GCAGAGAGCCATCGGCACCGTGTGATCGGCCCGCCCGCTGAACATCATCGGCGCA 637
Db 421 GCCAAGAGGCCATCGGCACCGTGTGTTGGGCGCCCGCCCGCTGAACATCATCGGCGCA 480
```

QY 638 ACATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCCCATCAGCCCATCGAGACCGTGC 697  
DB 481 ACCTGCTGACCGATCGGCTGCACCTTGAACCTTCCCCCATCAGCCCATCGAGACCGTGC 540  
QY 698 CCGTGAAGCTGAAGCCCGGATGGAACGCGCCCAAGGTGAAGCAAGTGGCCCTTGAACGAGG 757  
DB 541 CCGTGAAGCTGAAGCCCGGATGGAACGCGCCCAAGGTGAAGCAAGTGGCCCTTGAACGAGG 600  
QY 758 AGAAGATCAAGGCTGACCGCCATCTGCAGGAGATGGAGAGAGGAGGCAAGATCACCA 817  
DB 601 AGAAGATCAAGGCTGCTGGAGATCTGCACCGAGATGGAGAGAGGAGGCAAGATCACCA 660  
QY 818 AGATCGGCCCCGAGAACCCCTTACAACACCCCGCTTTCGCCCATCAAGAAAGAGACAGCA 877  
DB 661 AGATCGGCCCCGAGAACCCCTTACAACACCCCGCTTTCGCCCATCAAGAAAGAGACAGCA 720  
QY 878 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 937  
DB 721 CCAAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGG 780  
QY 938 AGGTGCAAGCTGGGATCCCCCAACCCCGCGGCTTGAAGAAAGAGAGGCTGACCGTGC 997  
DB 781 AGGTGCAAGCTGGGATCCCCCAACCCCGCGGCTTGAAGAAAGAGAGGCTGACCGTGC 840  
QY 998 TGGACGTGGCGAGCGCTTCTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG 1057  
DB 841 TGGACGTGGCGAGCGCTTCTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCG 900  
QY 1058 CCTTACCATCCCCAGATCAACAACGAGACCCCGGCTCCGCTACAGTACAAAGTGC 1117  
DB 901 CCTTACCATCCCCAGATCAACAACGAGACCCCGGCTCCGCTACAGTACAAAGTGC 960  
QY 1118 TGCCCAAGGCTGGAAGGAGAGCCCGAGATCTTTCAGAGCAGGATGACCAAGATCTGCG 1177  
DB 961 TGCCCAAGGCTGGAAGGAGAGCCCGAGATCTTTCAGAGCAGGATGACCAAGATCTGCG 1020  
QY 1178 AGCCCTTCCGCGCCGCAACCCCGAGATCTGTATCTACCA-----GGCCCGCCCTGTACG 1231  
DB 1021 AGCCCTTCCGCGAGCAGAACCCCGACATCTGTATCTACCAAGTACATGAGACGACTGTACG 1080  
QY 1232 TGGCAGCGACCTGGAGATCGGCAGCACCGCGCAAGATCGAGGAGCTCGCAAGCAC 1291  
DB 1081 TGGCAGCGACCTGGAGATCGGCAGCACCGCACCAAGATCGAGGAGCTCGCGAGCAC 1140  
QY 1292 TGCTGCGCTGGGTTTCAACACCCCGCAAGAAAGACACAGAGGAGCCCGCTTCTGT 1351  
DB 1141 TGCTGCGCTGGGTTTCAACACCCCGCAAGAAAGACACCAAGAGGAGCCCGCTTCTGT 1200  
QY 1352 GGATGGCTACGAGCTGCACCCCGACAAAGTGAACCGTGCAGCCCATCGAGCTGCGCGAGA 1411  
DB 1201 GGATGGCTACGAGCTGCACCCCGACAAAGTGAACCGTGCAGCCCATCATCTGCGCGAGA 1260  
QY 1412 AGGAGAGCTGGAACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCACGC 1471  
DB 1261 AGGACAGCTGGAACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCACGC 1320  
QY 1472 AGATCTAACCCGGGATCAAGGTGGCCAGCTGTGCAAGCTGTGCGCGGCGCCCAAGGCC 1531  
DB 1321 AGATCTAACCCGGGATCAAGGTGGAAGCAGCTGTGCAAGCTGTGCGCGGCGCCCAAGGCC 1380  
QY 1532 TGACCGACATCTGTCCTTCAACGAGGAGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1591  
DB 1381 TGACCGAGGTATCCCCCTGACCGAGGAGCGAGCTGGAGCTGGCGCGAGAACCGCGAGA 1440  
QY 1592 TCCTGCGGAGCCCGTGCAGCGGCTGTACTACGACCCAGCAAGGACCTGTGGCGCGAGA 1651  
DB 1441 TCCTGAAGGAGCCCGTGCACGAGTGTACTACGACCCAGCAAGGACCTGTGGCGCGAGA 1500  
QY 1652 TCCGAAGCAGGCGCACGACAGTGGACCTTACCGATCTTACAGGAGCCCTTCAAGAAC 1711  
DB 1501 TCCGAAGCAGGCGCACGAGGCGAGTGGACCTTACCGATCTTACAGGAGCCCTTCAAGAAC 1560

QY 1712 TGAAGACCGGCAAGTAGTAGCCCAAGATGCGCACCGCCCAACACCAACGACGTCGAGCAGTGA 1771  
DB 1561 TGAAGACCGGCAAGTAGTAGCCCGCGATGCGCGGCGCCCAACCAACGACGTCGAGCAGTGA 1620  
QY 1772 CCAGAGCGCTGCAAGAGATCGCCATGGAGAGCATGTGTATCTTGGGGCAAGACCCCCAAGT 1831  
DB 1621 CCAGAGCGCTGCAAGAGAGTGTAGCACCGAGAGCATGTGTATCTTGGGGCAAGATCCCCAAGT 1680  
QY 1832 TCCGCTTCCCATCCAGAAAGGAGACCTTGGGAGACCTTGGTGACCGACTACTTGGCAGGCCA 1891  
DB 1681 TCAAGCTTGCCCATCCAGAAAGGAGACCTTGGGAGCGCTTGTGTATGTAGTACTTGGCAGGCCA 1740  
QY 1892 CTTGGATCCCCCGAGTGGGAGTCTGTGAACAACCCCCCCCCCTTGGTGAAGCTGTGTGTACGAG 1951  
DB 1741 CTTGGATCCCCCGAGTGGGAGTCTGTGAACAACCCCCCCCCCTTGGTGAAGCTGTGTGTACGAG 1800  
QY 1952 TGAGAAAGGAGCCATCATCGGCGCGAGACCTTCTTAAGTGTGAGACGCGCGCCGCAACCGCG 2011  
DB 1801 TGAGAAAGGAGCCCATCTGTGGCGCGAGACCTTCTTAAGTGTGAGACGCGCGCCGCAACCGCG 1860  
QY 2012 AGACCAAGATCGGCAAGGCGGCTAGTGTAGCCGACCGGCGCGGAGCAAGATCGTGAGCC 2071  
DB 1861 AGACCAAGCTGGGCAAGGCGGCTAGTGTAGCCGACCGGCGCGGAGAGTGTGTAGCA 1920  
QY 2072 TGACCGAGACCAACAACAGAGACCGAGCTGCAGGCGCATCCAGCTGCGCCCTGCAGGACA 2131  
DB 1921 TCGCCGACACCAACAACAGAGACCGAGCTGCAGGCGCATCCACCTGCGCCCTGCAGGACA 1980  
QY 2132 GCGGAGCGAGGTGAACATCGTGACCGACAGCAGTACGCCCTTGGGATCATTCAGGCCC 2191  
DB 1981 GCGGCTTGGAGGTGAACATCGTGACCGACAGCAGTACGCCCTTGGGATCATTCAGGCCC 2040  
QY 2192 AGCCGACAGAGCAGGAGCGAGCTGTGAACCAAGATCATCGAGCAGCTCATCAAGAAGG 2251  
DB 2041 AGCCGACAGAGCAGGAGCGAGCTGTGAGCCAGATCATTCAGCAGCTCATCAAGAAGG 2100  
QY 2252 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGCGCATCGGCGGCAACGAGCAGATCG 2311  
DB 2101 AGAAGGTGTACCTGGCTGGGTGCCCGCCCAAGGCGCATCGGCGGCAACGAGCAGGTGG 2160  
QY 2312 ACAAGCTGTGAGCAAGGCGATCCGCAAGTGTCTTCTTGGAGCGGATCGATGGCGGCA 2371  
DB 2161 ACAAGCTGTGAGCGCGGCGATCCGCAAGTGTCTTCTTGAACGCGCATCGATGGCGGCA 2220  
QY 2372 TCGTGTATCTACCAAGTACATGAGCAGCTGTACGTGGGCGCGGCGCTAGGATCGATT 2431  
DB 2221 TCGTGTATCTACCAAGTACATGAGCAGCTGTACGTGGGCGCGGCGCTAGGATCGATT 2280  
QY 2432 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2463  
DB 2281 AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC 2312

RESULT 11  
BD263705  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BD263705 2300 bp DNA linear PAT 17-JUL-2003  
Improved expression of HIV polypeptides and production of virus-like particles.  
BD263705  
BD263705.1 GI:33073473  
JP 2002533124-A/72.  
synthetic construct  
synthetic construct  
other sequences; artificial sequences.  
1 (bases 1 to 2300)  
Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.  
Improved expression of HIV polypeptides and production of virus-like particles  
Patent: JP 2002533124-A 72 08-OCT-2002;  
CHIRON CORP  
OS Artificial Sequence  
PN JP 2002533124-A/72



Db 1795 AGGAGCCCATCTGTGGCGCCGAGACCTTCTACGTGACGCGCGCCCAACCGGAGACCA 1854

Qy 2018 AGATCGCAAGGCGGCTAGTGAACCGACCGGGCGGCGAGAGATCGTGAAGCTGACCG 2077

Db 1855 AGCTGGCAAGGCGGCTAGTGAACCGACCGGGCGGCGAGAGTGGTGAAGATCGCCG 1914

Qy 2078 AGACCAACCAAGGAGGAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCA 2137

Db 1915 ACACCACCAAGGAGGAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 1974

Qy 2138 GCGAGGTGAACATGTTGAACGAGCAGCAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2197

Db 1975 TGGAGGTGAACATGTTGAACGAGCAGCAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2034

Qy 2198 ACAGAGCGAGGAGGAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2257

Db 2035 ACAAGAGCGAGGAGGAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2094

Qy 2258 TGTACCTGAGTGTGGTGTGGCGCCCAAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2317

Db 2095 TGTACCTGGCTGGTGTGGCGCCCAAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2154

Qy 2318 TGTGAGCAAGGCGCATCCGCAAGTGTGTTCTTGAACGCGGATCGATGGCGGATCGTGA 2377

Db 2155 TGTGAGCGCGGCATCCGCAAGTGTGTTCTTGAACGCGGATCGATGGCGGATCGTGA 2214

Qy 2378 TCTACCAAGTACAGGAGGAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2437

Db 2215 TCTACCAAGTACAGGAGGAGTGTGAGGCGCATCCAGCTGGGCCCTGCAGGACAGCGGCC 2274

Qy 2438 TTCCCGGGGTAGCACCGGTGAATTC 2463

Db 2275 TTCCCGGGGTAGCACCGGTGAATTC 2300

RESULT 12

CO870575

LOCUS CO870575 2300 bp DNA linear PAT 13-SEP-2004

DEFINITION Sequence 83 from Patent EP1433851.

ACCESSION CO870575

VERSION CO870575.1 GI:52000091

KEYWORDS synthetic construct

SOURCE other sequences; artificial sequences.

ORGANISM

REFERENCE 1

AUTHORS Barnett,S., Zurmegede,J., Srivastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.

TITLE Improved expression of HIV polypeptides and production of virus-like particles

JOURNAL Patent: EP 1433851-A 83 30-JUN-2004;

CHIRON CORPORATION (US)

FEATURES

source

1. 2300

/organism="synthetic construct"

/mol\_type="unassigned DNA"

/db\_xref="taxon:32630"

/note="Description of Artificial Sequence:

FS(-).protmod.Rtopt.YMMW"

ORIGIN

Query Match 82.0%; Score 2019.2; DB 6; Length 2300;

Best Local Similarity 93.2%; Pred. No. 7.1e-202;

Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy 170 GCGCAAGGAGGCGCCACAGATGAAGACTGCAACGAGCGCGCAAGCCAACTTCTTCGCG 229

Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGAGACAGGCTAAATTCTTCGCG 60

Qy 230 AGGACCTTGGCTTCCCGGAGGCGCGCGGAGTTCCTCCGAGGACAGCAACCGGCGCA 289

Db 61 AGGACCTTGGCTTCCCGGAGGCGCGCGGAGTTCCTCCGAGGACAGCAACCGGCGCA 120

Qy 290 ACAGCCCCACACCGCCGAGCTGCAAGTGTGCGCGCG-----ACAAACCCCGCAGCGAGG 343

Db 121 ACAGCCCCACACCGCCGAGCTGCAAGTGTGCGCGCGCGAGAAACAAACAGCTTGAGCGAGG 180

Qy 344 CCGGCGCGAGGCGCGAGGCGACCTG-----NACTTCCCCCAGATCACCTGTGGCAGC 397

Db 181 CCGGCGCGAGGCGCGAGGCGACCTGAGCTTAACTTCCCCCAGATCACCTGTGGCAGC 240

Qy 398 GCCCCTTGGTGTGACATCAAGTGTGGCGCGCAGATCAAGGAGGCGCTGTCTGACACCGCGC 457

Db 241 GCCCCTTGGTGTGACATCAAGTGTGGCGCGCAGTCAAGGAGGCGCTGTCTGACACCGCGC 300

Qy 458 CCGACGACACCGTGTGTGGAGGAGATGAGCTTCCGCGCAAGTGTGAAAGCCCAAGATGATCG 517

Db 301 CCGACGACACCGTGTGTGGAGGAGTGAACCTGCGCGCAAGTGTGAAAGCCCAAGATGATCG 360

Qy 518 GCGGATCGCGCGCTTCAATCAAGTGTGGCGCGCAGTACGACCGAGATCTGTGATCTGCG 577

Db 361 GCGGATCGCGCGCTTCAATCAAGTGTGGCGCGCAGTACGACCGAGATCTGTGATCTGCG 420

Qy 578 GCAAGAGGCGCATCGGCACCGTGTGATCGCGCGCCACCCCGGTGAACATCATCGGCGCA 637

Db 421 GCCAAGAGGCGCATCGGCACCGTGTGTGGCGCGCCACCCCGGTGAACATCATCGGCGCA 480

Qy 638 ACATGTGTGACCCAGCTGGGCTGCACCTTGAATTTCCCATCAGCCCCATCGAGACCGTGC 697

Db 481 ACCTGTGTGACCCAGCTGGGCTGCACCTTGAATTTCCCATCAGCCCCATCGAGACCGTGC 540

Qy 698 CCGTGAAGCTGAAGCCCGGATGAGCGCGCCCAAGGTCAAGCAGTGGCGCTTGAACCGAGG 757

Db 541 CCGTGAAGCTGAAGCCCGGATGAGCGCGCCCAAGGTCAAGCAGTGGCGCTTGAACCGAGG 600

Qy 758 AGAAGATCAAGGCGCTGACCGCATCTGTGGAGGATGGAAGAGGAGGCGCAAGATCACCA 817

Db 601 AGAAGATCAAGGCGCTTGTGGAGATCTGTGGAGGATGGAAGAGGAGGCGCAAGATCACCA 660

Qy 818 AGATCGGCGCGGAGAACCCCTCAACACACCCCGTGTTCGCCATCAAGAAGAAGACAGCA 877

Db 661 AGATCGGCGCGGAGAACCCCTCAACACACCCCGTGTTCGCCATCAAGAAGAAGACAGCA 720

Qy 878 CCAAGTGTGGCGCAAGCTGTGGACTTTCGGGAGCTGAAACAGCGCGCACCCAGGACTTTCGG 937

Db 721 CCAAGTGTGGCGCAAGCTGTGGACTTTCGGGAGCTGAAACAGCGCGCACCCAGGACTTTCGG 780

Qy 938 AGGTGCAAGTGTGGCATCCCCACCCCGCGCTGAAAGAAAGAGAGCGTCAACCGTGC 997

Db 781 AGGTGCAAGTGTGGCATCCCCACCCCGCGCTGAAAGAAAGAGAGAGCGTCAACCGTGC 840

Qy 998 TGGACGTGGCGCAGCGCTACTTCAAGTGTGCGCTTGAACGAGGACTTTCGCAAGTACACCG 1057

Db 841 TGGACGTGGCGCAGCGCTACTTCAAGTGTGCGCTTGAACGAGGACTTTCGCAAGTACACCG 900

Qy 1058 CTTTACCATCCCCCAGCATCAACAGAGACCCCGCGCATCCGCTACCATCAACCGTGC 1117

Db 901 CTTTACCATCCCCCAGCATCAACAGAGACCCCGCGCATCCGCTACCATCAACCGTGC 960

Qy 1118 TGCCCGAGGCTGGAAGGCGAGCGCCAGCATCTTTCAGAGCAGCATCAACCAAGATCTCTGG 1177

Db 961 TGCCCGAGGCTGGAAGGCGAGCGCCAGCATCTTTCAGAGCAGCATCAACCAAGATCTCTGG 1020

Qy 1178 AGCCCTTTCGCGCGCGCAACCCCGAGATCTGTATCTACAGGCGCCCGCTGTAGTGGGCA 1237

Db 1021 AGCCCTTTCGCGCGCGCAACCCCGAGATCTGTATCTACAGGCGCCCGCTGTAGTGGGCA 1080

Qy 1238 GCGACCTTGGAGTTCGGCGCAGCGCGCGCAAGATCGAGGAGCTTGGCGAGCAGCTGCTGC 1297

Db 1081 GCGACCTTGGAGTTCGGCGCAGCGCGCGCAAGATCGAGGAGCTTGGCGAGCAGCTGCTGC 1140

Qy 1298 GCTTGGGCTTTCACACCCCGCAAGAGCAACAGAGGAGCGCCCGCTTCTGTGTGATGG 1357

Db 1141 GCTTGGGCTTTCACACCCCGCAAGAGCAACAGAGGAGCGCCCGCTTCTGTGTGATGG 1198

Qy 1358 GCTACGAGCTTGCACCCCGCAAGTGGAGCGGTGACGCCCATCTGAGCTGCGCGAGAGAGGA 1417

Db 1199 -----CGAGCTGCACCCGACAAAGTGGACCGTGCAGCCCATCATCTGCTCCGAGAAAGACA 1254  
Qy 1418 GCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGCCAGCCAGATCT 1477  
Db 1255 GCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAACTGGGCGCCAGCCAGATCT 1314  
Qy 1478 ACCCGGGGATCAAGGTGGCCAGCTGTGTGCAAGCTGTGCGGGGCGCCAGAGCCCTGACCG 1537  
Db 1315 ACGCCGGGATCAAGGTGAAGCAGCTGTGTGCAAGCTGTGCGGGGCGCCAGAGCCCTGACCG 1374  
Qy 1538 ACATGTGCTCCCTCACCGAGGAGCGGAGCTGTGAGCTGGCCGAGAAACCGGAGATCTGC 1597  
Db 1375 AGTGATCCCCCTACCGAGGAGCGGAGCTGTGAGCTGGCCGAGAACCGGAGATCTGGA 1434  
Qy 1598 GCAGACCGGTGCACGGCGGTGTACTACGACCCAGCAAGGACTGTGTGGCGCCAGATCCAGA 1657  
Db 1435 AGGAGCCCGTGCACGAGGTGTACTACGACCCAGCAAGGACTGTGTGGCGGAGATCCAGA 1494  
Qy 1658 AGCAGGCGCAGCACAGTGTGACCTACAGATCTTACAGGAGCCCTTCAAGAACTGTGAAGA 1717  
Db 1495 AGCAGGCGCAGGCGCAGTGACCTACAGATCTTACAGGAGCCCTTCAAGAACTGTGAAGA 1554  
Qy 1718 CCGCAGATGACCGCAAGTGGCACCGCCACACCAAGCAGCTGAAGCTGACCGAG 1777  
Db 1555 CCGCAAGTACGCCCGCATGCGCGCGCCACACCAAGCAGCTGAAGCAGCTGACCGAGG 1614  
Qy 1778 CCGTGCAGAAAGTCCCATGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCC 1837  
Db 1615 CCGTGCAGAAAGTGCAGCACGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1674  
Qy 1838 TGCCCATCCAGAAAGGAGACTGTGGAGACTGTGTGGACCGACTACTTGGCAGGCCACCTGGA 1897  
Db 1675 TGCCCATCCAGAAAGGAGACTGTGGAGCGCTGTGTGATGTAGTACTTGGCAGGCCACCTGGA 1734  
Qy 1898 TCCCGAGTGGAGTTGTGAAACACCCGCCCTTGTGTGAGCTGTGTGATGTGTAACAGCTGAGA 1957  
Db 1735 TCCCGAGTGGAGTTGTGAAACACCCGCCCTTGTGTGAGCTGTGTGATGTGTAACAGCTGAGA 1794  
Qy 1958 AGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGGCGCCGCAACCGCGAGACCA 2017  
Db 1795 AGGAGCCCATCTGTGGCGCGAGACCTTCTACGTGGACGGCGCCGCAACCGCGAGACCA 1854  
Qy 2018 AGATCGGCAAGCGCGCTAGTGACCGACCGGGGCGCGGAGAGATCGTGAGCTGTGACCG 2077  
Db 1855 AGCTGGGCAAGCGCGCTAGTGACCGACCGGGGCGCGGAGAGCTGTGTGAGCATCGCG 1914  
Qy 2078 AGACCAACCAAGAACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGCA 2137  
Db 1915 ACACCAACCAAGAACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC 1974  
Qy 2138 GCGAGGTGAACATCTGTACCGACAGCAGTACGCGCTTGGGCTCATCCAGGCCAGCCCG 2197  
Db 1975 TGGAGGTGAACATCTGTACCGACAGCAGTACGCGCTTGGGCTCATCCAGGCCAGCCCG 2034  
Qy 2198 ACAGAGCGAGAGCGAGCTGTGTGAACAGATCATCTGAGCAGCTGATCAAGAGGAGAGG 2257  
Db 2035 ACAGAGCGAGAGCGAGCTGTGTGAACAGATCATCTGAGCAGCTGATCAAGAGGAGAGG 2094  
Qy 2258 TGTAACCTGAGCTGGGTGCGCGCCCAAGGAGCATCGGGGCGCAAGCAGATCGAAGC 2317  
Db 2095 TGTAACCTGGGTGCGCGCCCAAGGAGCATCGGGGCGCAAGCAGATCGAAGC 2154  
Qy 2318 TGGTGAACAGGGGATCCGCAAGGTGTGTCTTGGACGGCATGATGAGGGGATCTGGA 2377  
Db 2155 TGGTGAACGGGATCCGCAAGGTGTGTCTTGGACGGCATGATGAGGGGATCTGGA 2214  
Qy 2378 TCTACAGTACATGGAGCAGCTGTACGTGGGACGCGGCGCTTAGGATCGATTAAAGC 2437  
Db 2215 TCTACAGTACATGGAGCAGCTGTACGTGGGACGCGGCGCTTAGGATCGATTAAAGC 2274  
Qy 2438 TTCCCGGGGTAGCACCGGTGAATTC 2463  
|||||

Db 2275 TTCCCGGGGTAGCACCGGTGAATTC 2300  
RESULT 13  
AR373388  
LOCUS AR373388 2300 bp DNA linear PAT 18-DEC-2003  
DEFINITION Sequence 83 from patent US 6602705.  
ACCESSION AR373388  
VERSION AR373388.1 GI:40075491  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2300)  
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.  
TITLE Expression of HIV polypeptides and production of virus-like  
particles  
JOURNAL Patent: US 6602705-A 83 05-AUG-2003;  
FEATURES  
source location/Qualifiers  
1..2300  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 82.0%; Score 2019.2; DB 6; Length 2300;  
Best Local Similarity 93.2%; Pred. No. 7.1e-202;  
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;  
Qy 170 GCGCAAGGAGGGCCACAGATGAAGACTGCACCGAGCGCCAGCCAACTTCTTCGCG 229  
Db 1 GCGGCCGCGAAGGACCAAAATGAAGATTGCATGAGACAGGTAATTTCTTCGCG 60  
Qy 230 AGGACCTTGGCTTCTCCCGCAGGCAAGCCGCGAGTTTCCCGAGCAGCAGAACCGCGCA 289  
Db 61 AGNACTTGGCTTCTTCAGGGCAAGCCGCGAGTTTTCAGCAGCAGCAGAACCGCGCA 120  
Qy 290 ACAGCCCAACAGCCGCGAGCTGCAGTGTGCGCGCG-----ACAACCCCGCAGCGAGG 343  
Db 121 ACAGCCCAACCGCGCGAGCTGCAGTGTGCGCGCGAGAAACAACAGCCTGAGCGAGG 180  
Qy 344 CCGCGCGCAGCGCGCGAGCGCACCTG-----AACTTCCCGCAGATCACCTGTGGCAGC 397  
Db 181 CCGCGCGCAGCGCGCGAGCGCACCTGAGCTTCAACTTCCCGCAGATCACCTGTGGCAGC 240  
Qy 398 GCCCTCTGTGTAGCATCAAGGTGGCGCGCCAGATCAAGAGGCGCTGTGTGACACCGCGC 457  
Db 241 GCCCTCTGTGTAGCATCAGGATCGCGCGCGCAGCTCAAGAGGCGCTGTGTGACACCGCGC 300  
Qy 458 CCGACGACACCGTGTGTGAGGAGATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCG 517  
Db 301 CCGACGACACCGTGTGTGAGGAGATGAACCTGCGCGCGCAAGTGAAGCCCAAGATGATCG 360  
Qy 518 GCGGCATCGCGGGCTTCATCAAGTGTGCGCCAGTACGACAGATCCTGATCGAGATCTGCG 577  
Db 361 GCGGCATCGCGGGCTTCATCAAGTGTGCGCGAGTACGACAGATCCTCCGTGGAGATCTGCG 420  
Qy 578 GCAAGAGGCGCATCGCGCACCGTGTGTGATCGCGCGCCACCGCGTGAACATCATCGCGCGCA 637  
Db 421 GCAAGAGGCGCATCGCGCACCGTGTGTGCGCGCCACCGCGTGAACATCATCGCGCGCA 480  
Qy 638 ACATGTGACCCAGCTGGGTGACCCCTGAACTTCCCATCAGCCCATCGAGACCGTGC 697  
Db 481 ACCTGTGACCCAGATCGGCTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGC 540  
Qy 698 CCGTGAAGTGAAGCCCGGCATGGAACCGCCCAAGTGAAGCAGTGGCCCTGACCGAGG 757  
Db 541 CCGTGAAGTGAAGCCCGGCATGGAACCGCCCAAGTGAAGCAGTGGCCCTGACCGAGG 600  
Qy 758 AGAAGATCAAGGGCCCTGACCGCATCTGCGAGGAGATGGAAGAGGAGGCAAGATCACCA 817  
Db 601 AGAAGATCAAGGGCCCTGCTGTGAGATCTGCACCGGATGGAAGAGGAGGCAAGATCAGCA 660  
Qy 818 AGATCGGCGCGCGAGAACCCCTTCAACAACCCCGGTGTTGCCATCAAGAAAGAGCAGCA 877

```
|||||
661 AGATCGGCCCCGAGACCCCTTCAACACACCCCGCTGTTCGCCATCAAGAGAAGGACGCA 720
QY
878 CCAAGTGGCGCAAGCTGGTGGATTTTCGCGAGCTGAAACAAAGCGCACCCAGGATCTTCGGG 937
Db
721 CCAAGTGGCGCAAGCTGGTGGATTTTCGCGAGCTGAAACAAAGCGCACCCAGGATCTTCGGG 780
QY
938 AGGTGAGCTGGGATCCCCACCCCGCGCGCTGAGAGAGAGAGAGGCTGACCGTGC 997
Db
781 AGGTGAGCTGGGATCCCCACCCCGCGCGCTGAGAGAGAGAGAGGCTGACCGTGC 840
QY
998 TGGACGTGGGCGACGCTACTTTCAGGGTGCCTTCGACGAGGACTTTCGCAAGTACACCG 1057
Db
841 TGGACGTGGGCGACGCTACTTTCAGGGTGCCTTCGACGAGGACTTTCGCAAGTACACCG 900
QY
1058 CTTTCAACATCCCGAGATCAACAAAGAGACCCCGCGCATTCGCTACCAAGTACAAAGTGC 1117
Db
901 CTTTCAACATCCCGAGATCAACAAAGAGACCCCGCGCATTCGCTACCAAGTACAAAGTGC 960
QY
1118 TGGCCCGAGGGCTGGAGGGCAGCCCGAGATCGTGATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177
Db
961 TGGCCCGAGGGCTGGAGGGCAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020
QY
1178 AGCCCTTTCGCGCGCCGCAACCCCGAGATCGTGATCTTACCAGGCCCCCTTCTAGCTGGGCA 1237
Db
1021 AGCCCTTTCGCGAGACGACCCCGCATCTGTGATCTTACCAGGCCCCCTTCTAGCTGGGCA 1080
QY
1238 GCGACTGGAGATTCGCGCAGCACCGCGCCCAAGATTCGAGGAGCTGCGCAAGCACCTGCTGC 1297
Db
1081 GCGACTGGAGATTCGCGCAGCACCGCACCAAGATTCGAGGAGCTGCGCAGCACCTGCTGC 1140
QY
1298 GCTGGGGCTTACACACCCCGCAAGAGACACAGAGAGGAGCCCTTCTCTGTGATGG 1357
Db
1141 GCTGGGGCTTACACACCCCGCAAGAGACACAGAGAGGAGCCCTTCTCTGTGATGG 1198
QY
1358 GCTTACAGCTGACACCCCGCAAGTGAACCGTGCAGGCCCATTCGAGCTGCCCGAGAGGAGA 1417
Db
1199 -----CGAGCTGCACCCCGCAAGTGAACCGTGCAGGCCCATTCGCTGCCCGAGAGGAGA 1254
QY
1418 GCTGGACCGTGAACGACATCCAGAACGCTGTGGGCAAGCTGAACTGGGCGCAGCGAGATCT 1477
Db
1255 GCTGGACCGTGAACGACATCCAGAACGCTGTGGGCAAGCTGAACTGGGCGCAGCGAGATCT 1314
QY
1478 ACCCGCGCATCAAGTGGCGGACCTGTGAAGCTGTGCGGCGCCAGGCGCCCTGACCG 1537
Db
1315 AC CGCGCATCAAGGTGAAGCAGCTGTGAAGCTGTGCGGCGCACCAAGGCCCTGACCG 1374
QY
1538 ACATCGTGCCTGACCGAGGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCCTGC 1597
Db
1375 AGGTGATCCCTTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCCTGA 1434
QY
1598 GCGAGCCCGTGCACGCGGTGTAACGACCCCGAGAGGACCTGTGGCGGAGATCCAGA 1657
Db
1435 AGGAGCCCGTGCACGAGGTGTAACGACCCCGAGAGGACCTGTGGCGGAGATCCAGA 1494
QY
1658 AGCAGGGCCACGACCACTGACATCCAGATCTTACCAGGAGCCCTTCAAGAACCTGAGA 1717
Db
1495 AGCAGGGCCACGAGGACCTGAGACCTTACAGATCTTACCAGGAGCCCTTCAAGAACCTGAAGA 1554
QY
1718 CCGGCAAGTACGCGCATCGCGGCGCCACACCAACGACGCTGAAGCAGCTGACCGAGG 1777
Db
1555 CCGGCAAGTACGCGCATCGCGGCGCCACACCAACGACGCTGAAGCAGCTGACCGAGG 1614
QY
1778 CCGTGCAGAGATCGGCATGGAGAGCATCTGTGATCTTGGGCGAAGACCCCAAGTTCGCGC 1837
Db
1615 CCGTGCAGAGATCGGACCGAGAGATCTGTGATCTTGGGCGAAGATCCCAAGTTCAGG 1674
QY
1838 TGGCCCATCCAGAGGAGACCTGGGAGACCTGTGGACCGACTACTGCGAGGCGCACCTGGA 1897
Db
1675 TGGCCCATCCAGAGGAGACCTGGGAGGCTGTGGTGGATGGAGTACTGCGAGGCGCACCTGGA 1734
QY
1898 TCCCGGAGTGGGAGTTTCGTGAACACACCCCGCTGTGTGAAGCTGTGTGATCAGCTGGAGA 1957
```

```
Db
1735 TCCCCGAGTGGGAGTTTCGTGAACACACCCCGCTGTGTGAAGCTGTGTACCAAGCTGGAGA 1794
QY
1958 AGGAGCCCATCATCGGCGCGGAGACCTTCTAGTGAACGCGCGCCCAACCCGCGAGACCA 2017
Db
1795 AGGAGCCCATCGTGGCGCGGAGACCTTCTAGTGAACGCGCGCCCAACCCGCGAGACCA 1854
QY
2018 AGATCGGCAAGGCGCGCTACGTGACCGAACCGGGCGCGGAGAAAGATCGTGAAGCTCGACCG 2077
Db
1855 AGCTGGGCAAGCGCGCTACGTGACCGAACCGGGCGCGGAGAGGTGTGTGAGCATCGCG 1914
QY
2078 AGACCAACCAACCGAAGACCGAGCTGCAAGCCATTCAGCTGGCCCTGCAAGGACAGCGGCA 2137
Db
1915 ACACCAACCAACCGAAGACCGAGCTGCAAGCCATTCACCTGGCCCTGCAAGGACAGCGGCC 1974
QY
2138 GCGAGGTGAACATCGTGAACGAGCAGTACCGCTTGGGCATCATCCAGGCCACGCGCG 2197
Db
1975 TGGAGGTGAACATCGTGAACGAGCAGCAGTACCGCTTGGGCATCATCCAGGCCACGCGCG 2034
QY
2198 ACAAGAGCGAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAAGG 2257
Db
2035 ACNAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAGAGGAGAGG 2094
QY
2258 TGTACCTGAGCTGGTGGCGGCCCAAGGGGATTCGGCGGCAACGAGCAGATCGACAAGC 2317
Db
2095 TGTACCTGAGCTGGTGGCGGCCCAAGGGGATTCGGCGGCAACGAGCAGTGGACAAGC 2154
QY
2318 TGGTGAAGAGGCGATCCGCGAGTGTCTTCTGGAACGCGCATCGATGGCGCATCGTGA 2377
Db
2155 TGGTGAAGCGCGGATCCGCGAGTGTCTTCTGGAACGCGCATCGATGGCGCATCGTGA 2214
QY
2378 TCTTACAGTACATGGAACGACCTGTAGCTGGGCGAGCGCGGCCCTAGGATCGATTTAAAGC 2437
Db
2215 TCTTACAGTACATGGAACGACCTGTAGCTGGGCGAGCGCGGCCCTAGGATCGATTTAAAGC 2437
QY
2438 TCCCCGGGCTAGCACCGGTGAATTC 2463
Db
2275 TCCCCGGGCTAGCACCGGTGAATTC 2300

RESULT 14
AX427936
LOCUS
DEFINITION
Sequence 174 from Patent WO0232943.
ACCESSION
AX427936
VERSION
AX427936.1 GI:21538023
KEYWORDS
synthetic construct
SOURCE
synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1
AUTHORS
Huang, Y. and Nabel, G.J.
TITLE
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
JOURNAL
Patent: WO 0232943-A 174 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
Location/Qualifiers
source
1..9788
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 82.0%; Score 2019.2; DB 6; Length 9788;
Best Local Similarity 91.6%; Pred. No. 4.8e-202;
Matches 2190; Conservative 0; Mismatches 178; Indels 24; Gaps 4;

QY
14 TGGCCCGAGGCGCATGAGCCAGGCGCACCGAGGCGCCCAACATCTCTGATGCGAGCGCAGCAACTCA 73
Db
2961 TGGCCCGAGGCGCATGAGCAAGGTGAACACCAACCATCATCATGATGCGAGCGCAGCACTGCA 3020
QY
74 AGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCGAAGGAGGCCACATCGCCCGCA 133
```



Db 3021 AGGGCCCCAGCGCACCATCAAGTGTCTTCAACTGGGGCAGGAGGGCCACCTGGCCCCGCA 3080  
Qy 134 ACTCGCGCCCCCGCGAAGAGGGCTGCTGGAAAGTGGCGAAGAGGGCCACACAGATGA 193  
Db 3081 ACTCGCGCCCCCGCGAAGAGGGCTGCTGGAAAGTGGCGAAGAGGGCCACACAGATGA 3140  
Qy 194 AGGACTGCACGAGGGCCAGGCGCAACTTCTTCGCGGAGGACTGGCCCTTCCCCCAGGGCA 253  
Db 3141 AGGACTGCACGAGGGCGAGGCTTAA-----TAGGGGAAGATCTGGCCCTTCCCAAGGGGA 3194  
Qy 254 AGGCGCGGAGTTCCTCAGCGAGCAGAAACCGCGCAACAGCGCCACACAGCGCGCGAGTGC 313  
Db 3195 AGGCGAGGGAAATTTCTTCAGAGCAGACCAAGAGCAACAGGCCCAACAGAGAGAGCTTC 3254  
Qy 314 AGGTGCGCG-----CGACAACCCCGCAGCGAGGCGCGCGCGAGCGCCAGGGCA--- 364  
Db 3255 AGGTTTGGGGAAGAGACAACAACCTCCCTCTCAGAAGCAGGAGCGCCGATAGACAAGGAATG 3314  
Qy 365 ---CCCTGAACCTTCCCGCAGATCACCTCTGTGGCAGCGCCCTGCTGAGCATCAAGGTGG 421  
Db 3315 TATCTTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAAAGATAG 3374  
Qy 422 GCGGCCAGATCAAGAGAGCCCTGTGGACACCGCGCGAGCACACCGTGTCTGGAGGAGA 481  
Db 3375 GGGGCCAGCTGAGAGAGGCCCCCTTAGACACCGGCGCGAGCACACCGTGTCTGGAGGAGA 3434  
Qy 482 TGAGCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGCATCGGCGGCTTCAAGG 541  
Db 3435 TGAACCTGCCCGCGCTGGAAGCCCCAAGATGATCGCGCGCATCGGCGGCTTCAATCAAGG 3494  
Qy 542 TGCGCCAGTAGCACAGATCTCTGATCGAGATCTCTGGGCAAGAGGCCATCGGCAACCGTGC 601  
Db 3495 TGGGCCAGTAGCACAGATCTCTGATCGAGATCTCTGGGCAACAGGCCATCGGCAACCGTGC 3554  
Qy 602 TGATCGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTCACCCAGCTGGGCTGCA 661  
Db 3555 TGGTGGGCCCCACCCCGTGAACATCATCGGCGCAACATGCTCACCCAGATCGGCTGCA 3614  
Qy 662 CCCTGAATCTTCCCATCAGCCCCCATCGAGACCGTGGCCCGTGAAGTGAAGCCCGGCATGG 721  
Db 3615 CCCTGAATCTTCCCATCAGCCCCCATCGAGACCGTGGCCCGTGAAGTGAAGCCCGGCATGG 3674  
Qy 722 ACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCA 781  
Db 3675 ACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGCTGGAGA 3734  
Qy 782 TCTCGGAGGAGATGGAGAGAGGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACA 841  
Db 3735 TCTGCAACGAGATGGAGAGAGGGCAAGATCAGCAAGATCGGCCCGCGAGAACCCCTACA 3794  
Qy 842 ACACCCCGCTTTCGCCATCAAGAGAAGGAGACAGCACCAAGTGGCGCAAGCTGGTGACT 901  
Db 3795 ACACCCCGCTTTCGCCATCAAGAGAAGGAGACAGCACCAAGTGGCGCAAGCTGGTGACT 3854  
Qy 902 TCCGCGAGCTGAACAAGCGCAACCAAGGACTTCTGGGAGGTGACGTGGGCATCCCCCACC 961  
Db 3855 TCCGCGAGCTGAACAAGCGCAACCAAGGACTTCTGGGAGGTGACGTGGGCATCCCCCACC 3914  
Qy 962 CCGCGCGCTGAAGAAGAGAGCGTGAACCGTGTGGAGCTGGCGGCGCGCCCTACTTCA 1021  
Db 3915 CCGCGCGCTGAAGAAGAGAGCGTGAACCGTGTGGAGCTGGCGGCGCGCCCTACTTCA 3974  
Qy 1022 GCGTGCCTTGGACGAGGACTTTCGCAAGTAGTACACCGCTTCAACCATCCCCAGCATCAACA 1081  
Db 3975 GCGTGCCTTGGACAGGACTTTCGCAAGTAGTACACCGCTTCAACCATCCCCAGCATCAACA 4034  
Qy 1082 ACAGAGACCCCGCATCCGCTTACAGTACAAAGTGTGCCCCAGGGCTGGAAAGGGCAGCC 1141  
Db 4035 ACAGAGACCCCGCATCCGCTTACAGTACAAAGTGTGCCCCAGGGCTGGAAAGGGCAGCC 4094  
Qy 1142 CCAGCATCTTCCAGAGCAGCATACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCG 1201  
Db 4095 CCGCATCTTCCAGTGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCAAGCAGAACCCCG 4154

Qy 1202 AGATCTGTGATCTACCA-----GGCCCCCTGTGTAGTGGGAGGAGCACTTGAGATTCGGCC 1255  
Db 4155 ACATCTGTGATCTACCACTGTATGACCACTGTACGTGGGAGGAGCACTTGAGATTCGGCC 4214  
Qy 1256 AGCACCGCGCCAGATCGAGGAGCTGGCAAGCACTGTCTGCGTGGGGCTTCCACCAACC 1315  
Db 4215 AGCACCGCGCCAGATCGAGGAGCTGGCCAGCACTGTCTGCGTGGGGCTTCCACCAACC 4274  
Qy 1316 CCGACAAGAGACCAAGAGGAGCGCCCTTCTCTGTGGATGGGCTAGAGCTGCACCCCG 1375  
Db 4275 CCGACAAGAGACCAAGAGGAGCGCCCTTCTCTGTGGATGGGCTAGAGCTGCACCCCG 4334  
Qy 1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAAAGGAGAGTGTGAACCTGAACGACA 1435  
Db 4335 ACAAGTGGACCGTGCAGCCCATCGTGTGCGCGAGAAAGGAGAGTGTGAACCTGAACGACA 4394  
Qy 1436 TCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCGGCATCAAGTGC 1495  
Db 4395 TCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCGGCATCAAGTGC 4454  
Qy 1496 GCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGGCCCTTGACCGACATCGTGCCTGACCG 1555  
Db 4455 GCCAGCTGTGCAAGCTGCTGCGCGCGCCCAAGGCCCTTGACCGAGGTGTGCTGCTGACCG 4514  
Qy 1556 AGGAGCGCGAGCTGGAGCTGGCCGAGAAACCGCGAGATCTCTGCGGAGAGCCCTGTGACGGCG 1615  
Db 4515 AGGAGCGCGAGCTGGAGCTGGCCGAGAAACCGCGAGATCTCTGAAGGAGAGCCCTGTGACGGCG 4574  
Qy 1616 TGTACTACGACCCCGCAGCAAGACCTGTGTGGCGGAGATCTCAGAACGAGGGCCACACCAAGT 1675  
Db 4575 TGTACTACGACCCCGCAGCAAGACCTGTGTGGCGGAGATCTCAGAACGAGGGCCAGGGCCAGT 4634  
Qy 1676 GGACTTACAGATCTTACAGAGGCTTCAAGAACCTGAAGACCGCGAAGTACCGCAAGA 1735  
Db 4635 GGACTTACAGATCTTACAGAGGCTTCAAGAACCTGAAGAACCTGAAGACCGCGAAGTACCGCGCA 4694  
Qy 1736 TGCGCACCGGCCACACCAACGAGCTGAAGCAGCTGACCGAGGCGCGTGCAGAAAGATTCGCCA 1795  
Db 4695 TGAAGGGCGGCCACACCAACGAGCTGACCGAGCAGCTGACCGAGGCGCGTGCAGAAAGATTCGCCA 4754  
Qy 1796 TGGAGACATCTGTGATCTGGGCGAAGACCCCGAAGTTCGCGCTGCCATCCAGAGGAGA 1855  
Db 4755 CCGAGACATCTGTGATCTGGGCGAAGACCCCGAAGTTCGCGCTGCCATCCAGAGGAGA 4814  
Qy 1856 CTTGGGAGACTGTGTGACCGACTACTGTGGCAGGCGCACTGTGATCCCGAGTGGGAGTTCG 1915  
Db 4815 CTTGGGAGGCTGTGTGACCGACTACTGTGGCAGGCGCACTGTGATCCCGAGTGGGAGTTCG 4874  
Qy 1916 TGAACACCCCGCTGGTGAAGCTGTGGTACCAAGCTGGAGAGAGGCCCATCATCGGCG 1975  
Db 4875 TGAACACCCCGCTGGTGAAGCTGTGGTACCAAGCTGGAGAGAGGCCCATCATCGGCG 4934  
Qy 1976 CCGAGACTTCTAGCTGGAAGCGCGCCCGCAACCGCGAGACCAAGATCGGAGAGGCCCGCT 2035  
Db 4935 CCGAGACTTCTAGCTGGAAGCGCGCCCGCAACCGCGAGACCAAGCTGGGAGAGGCCCGCT 4994  
Qy 2036 ACGTGACCGAGCCGGGCGCGCAGAAAGATCGTGAGCTGACCGAGACCAACCAACAGAGAGA 2095  
Db 4995 ACGTGACCGAGCCGGGCGCGCAGAAAGTGTGTGCCCCCTGACCGACACCAACCAAGAGAGA 5054  
Qy 2096 CCGAGCTGCAAGGCCATCCAGCTGGCCCTGCAAGGACAGCGGCGAGGTGAACATCTGTA 2155  
Db 5055 CCGAGCTGCAAGGCCATCCACTGGCCCTGCAAGGACAGCGGCTGGAGGTGAACATCTGTA 5114  
Qy 2156 CCGAGCAGATGACCGCCCTGGGCGATCATCAGGCCCGCAGCGACCAAGAGGAGAGCGAGC 2215  
Db 5115 CCGAGCAGATGACCGCCCTGGGCGATCATCAGGCCCGCAGCGAGCGAGCGAGCGAGC 5174  
Qy 2216 TGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGC 2275  
Db 5175 TGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGC 5234

```
QY 2276 CCGCCACAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGGTGCAGCAAGGGCATCC 2335
Db 5235 CCGCCACAAGGGCATCGCGGCAACGAGCAGTGGACGGCTGGTGAGCGCGGCATCC 5294

QY 2336 GCAAGGTGCTGTTCTTGAGCGGCATCGATGGCGGCATCGTGATCTACCACTGA 2387
Db 5295 GCAAGGTGCTGTTCTTGAGCGGCATCGACAAAGGCCCGAGGAGGAGCAGAGAA 5346

RESULT 15
AX427931
LOCUS AX427931 9169 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 169 from Patent WO0232943.
ACCESSION AX427931
VERSION AX427931.1 GI:21538018
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Huang Y. and Nabel G.J.
AUTHORS Modifications of hiv env, gag, and pol enhance immunogenicity for
TITLE Genetic immunization
JOURNAL Patent: WO 0232943-A 169 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1. 9169
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="plasmid pVR1012x/s containing HIV genes"

ORIGIN

Query Match 81.4%; Score 2005.2; DB 6; Length 9169;
Best Local Similarity 91.4%; Pred. No. 1.4e-200;
Matches 2187; Conservative 0; Mismatches 183; Indels 22; Gaps 5;

QY 14 TGSCCAGGCGCATGAGCGGCCACAGCGGCCAACATCTTGATGCGAGCGGCAACTTCA 73
Db 2967 TGSCCAGGCGCATGAGCGGCCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3026

QY 74 AGGCCCCCAAGCGCATCATCAAGTGTCTCAATGCGCGCAAGGAGGCGCCACATCGCCGCA 133
Db 3027 AGGCCCCAAGCGCAT---CAAGTGTCTCAATGCGCGCAAGGAGGCGCCACCTGGCCGCA 3083

QY 134 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAACTGGCGCAAGGAGGCGCCACCATGA 193
Db 3084 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGGAACTGGCGCAAGGAGGCGCCACCATGA 3143

QY 194 AGGACTGCACCGAGCGCGCCAACTTCTTCGCGAGGACCTGGCTTCCCCCAGGGCA 253
Db 3144 AGGACTGCACCGAGCGAGCAGGCTTAA-TTTTTTAAAGGAGATCTGGCTTCCCAAGAGGA 3202

QY 254 AGGCGCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAAGCCCCACAGCGCGGAGCTGC 313
Db 3203 AGSCCAGGGAATTTCTTCAGAGCAGACAGAGCCCAAGAGAGGAGGCTTC 3262

QY 314 AGGTGGCGCG------CGAACCCCGCGAGCGCGCGCGCGCGCGCGCGCA--- 364
Db 3263 AGGTGGGGAAGAGAACAACTCCCTCTCAGAAAGCAGGAGCGGATAGAACAGGAACTG 3322

QY 365 ---CCCTGAACTTCCCCAGATCACTGTGGCAGCGCCCTTGTGAGCATCAAGGTGG 421
Db 3323 TATCTTTTAGCTTCCCTCAGATCACTTTTGGCAGCGACCCCTCGTCACATTAAGATAG 3382

QY 422 GCGGCCAGATCAAGAGGCGCTGTGAGACACCGCGCGCGCAGCAGACCGTGTGAGGAGGA 481
Db 3383 GGGGCCAGCTGAAGGAGGCGCTTTAGACACCGCGCGCGCAGCAGACCGTGTGAGGAGGA 3442

QY 482 TGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAGG 541
Db 3443 TGAACCTGCCCGCGCTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAGG 3502
```

```
QY 542 TGGCCGAGTACGACCAAGATCTTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGC 601
Db 3503 TGGCCGAGTACGACCAAGATCTTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGC 3562

QY 602 TGATCGGCGCCACCCCGGTGAACATCATCGGCGGCAACATGCTGAGCCAGCTGGGTGCA 661
Db 3563 TGGTGGGCGCCACCCCGGTGAACATCATCGGCGGCAACCTGCTGAGCCAGATCGGTGCA 3622

QY 662 CCTGAATCTTCCCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGATGG 721
Db 3623 CCTGAATCTTCCCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAAGCCCGGATGG 3682

QY 722 ACGGCCCCCAAGGTGAAGCAGTGCCTGACGAGGAGAGATCAAGGCCCTGACCGCCA 781
Db 3683 ACGGCCCCCAAGGTGAAGCAGTGCCTGACGAGGAGAGATCAAGGCCCTGGTGAGA 3742

QY 782 TTTGAGAGAGATGGAAGAGAGGGGCAAGATCAACAGATCGGCCCGGAGAACCCCTACA 841
Db 3743 TTTGCAACCGAGATGGAAGAGAGGGGCAAGATCAGCAAGATCGGCCCGGAGAACCCCTACA 3802

QY 842 ACACCCCGCTGTTCCGCATCAAGAGAGAGAGACCAAGTGGCGCAAGCTGCTGTGACT 901
Db 3803 ACACCCCGCTGTTCCGCATCAAGAGAGAGAGACCAAGTGGCGCAAGCTGCTGTGACT 3862

QY 902 TCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGCAAGCTGGGCATCCCCACC 961
Db 3863 TCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGAGGTGCAAGCTGGGCATCCCCACC 3922

QY 962 CGCGCGGCTGAAGAGAGAGAGCGTGAACCGTGTGACGTGGAGCGGAGCGCTTACTTCA 1021
Db 3923 CGCGCGGCTGAAGAGAGAGAGCGTGAACCGTGTGACGTGGAGCGGAGCGCTTACTTCA 3982

QY 1022 GGTGCGCTGGACGAGGACTTCCGCAAGTACACCCCTTTCACCATCCCGCAGCATCAACA 1081
Db 3983 GGTGCGCTGGACGAGGACTTCCGCAAGTACACCCCTTTCACCATCCCGCAGCATCAACA 4042

QY 1082 ACAGAGACCCCGGCGATCCGCTACCACTACAGTACAAAGTGTGCTGCCCAAGGCTGGAAGGGAGCC 1141
Db 4043 ACAGAGACCCCGGCGATCCGCTACCACTACAGTACAAAGTGTGCTGCCCAAGGCTGGAAGGGAGCC 4102

QY 1142 CMAGCATCTTCAGAGCAGATGACCAAGATCTTGAGCGCTTTCGCGGCCCGCAACCCCG 1201
Db 4103 CGGCCATCTTCCAGTGCAGCATGACCAAGATCTCGAGCGCTTTCGCAAGAGACGAAACCCCG 4162

QY 1202 AGATCGTGTATCTACCA- ----GGCCCCCTGTACGTGGCGCAGCAGCTTGGAGATCGGCC 1255
Db 4163 ACATCGTGTATCTACCAAGTACATGGAACACTGTACGTGGCGCAGCAGCTTGGAGATCGGCC 4222

QY 1256 AGCACCGCGCCAGATCGAGGAGCTGCGCAAGCACTGTGTGCTGGGGCTTCAACACCC 1315
Db 4223 AGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGTGTGCTGGGGCTTCAACACCC 4282

QY 1316 CCGAAGAAGCAACGAGAGGAGCCCTTCTGTGTGATGGGCTAGGAGTGCACCCCG 1375
Db 4283 CCGAAGAAGCAACGAGAGGAGCCCTTCTGTGTGATGGGCTAGGAGTGCACCCCG 4342

QY 1376 ACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGAACGTGAACGACA 1435
Db 4343 ACAAGTGGACCGTGCAGCCCATCGTGTGCCGAGAGAGAGAGCTGGAACGTGAACGACA 4402

QY 1436 TCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCCAGATCTTACCCCGGATCAAGGTGC 1495
Db 4403 TCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCGCAGCCAGATCTTACCCCGGATCAAGGTGC 4462

QY 1496 GCAGCTGTGCAAGCTGTGCGCGGCGCCAGGCGCTTGACCCGACATGTCGCCCTGACCG 1555
Db 4463 GCAGCTGTGCAAGCTGTGCGCGGCGCCAGGCGCTTGACCCGAGGTGGTGGCTGACCG 4522

QY 1556 AGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGGAGCCCGTGCAGCGCG 1615
Db 4523 AGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTTGGAGGAGCCCGTGCAGCGCG 4582

QY 1616 TGTACTACGACCCCGCAGCAAGGACCTGTTGGTGGCGAGATCCAGAGAGCGGCGCACGACAGT 1675
```

[illegible]

Search completed: June 1, 2005, 20:20:37  
Job time : 10550.7 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 09:47:50 ; Search time 389.948 Seconds  
(without alignments)  
10360.282 Million cell updates/sec

Title: US-09-610-313B-30  
Perfect score: 2469  
Sequence: 1 gtcagcgcaccatgcccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2052	83.1	2312	4	US-09-475-515-84
2	2025.2	82.0	2306	4	US-09-475-515-82
3	1998.4	80.9	2300	4	US-09-475-515-83
4	1969.4	79.8	4319	4	US-09-475-515-6
5	1858	75.3	2305	4	US-09-475-515-80
6	1831.2	74.2	2299	4	US-09-475-515-81
7	1678.6	68.0	4307	4	US-09-552-950-2
8	1659.4	67.2	4307	4	US-09-936-572-2
9	1651.4	66.9	4327	4	US-09-936-572-14
10	1651.4	66.9	4353	4	US-09-936-572-13
11	1651.4	66.9	4642	4	US-09-936-572-12
12	1651.4	66.9	9772	4	US-09-552-950-5
13	1592.8	64.5	8366	4	US-09-872-733A-6
14	1587.4	63.1	4338	4	US-09-872-733A-1
15	1336.4	54.1	2577	4	US-09-952-060-1
16	1336.4	54.1	2650	4	US-09-952-060-5
17	1327.4	53.8	4053	4	US-09-952-060-34
18	1325.2	53.7	2577	4	US-09-952-060-3
19	1325.2	53.7	2650	4	US-09-952-060-7
20	1325.2	53.7	38519	4	US-09-952-060-28
21	1227.2	49.7	9010	4	US-09-184-418C-8
22	1199.4	48.6	9913	4	US-09-827-688-11
23	1196.2	48.4	8972	4	US-09-184-418C-9
24	1190.6	48.2	2467	4	US-09-872-733A-3
25	1189	48.2	8959	4	US-09-184-418C-11
26	1165.6	47.2	8992	4	US-09-184-418C-4
27	1132	45.8	2601	3	US-09-117-217-7

28 1132 45.8 2601 3 US-09-117-217-9 Sequence 9, Appli  
29 1132 45.8 2601 3 US-09-117-217-11 Sequence 11, Appli  
30 1132 45.8 2601 3 US-09-117-217-13 Sequence 13, Appli  
31 1132 45.8 2601 4 US-09-735-487-7 Sequence 7, Appli  
32 1132 45.8 2601 4 US-09-735-487-9 Sequence 9, Appli  
33 1132 45.8 2601 4 US-09-735-487-11 Sequence 11, Appli  
34 1132 45.8 2601 4 US-09-735-487-13 Sequence 13, Appli  
35 1132 45.8 4307 4 US-09-552-950-1 Sequence 1, Appli  
36 1132 45.8 4307 4 US-09-936-572-1 Sequence 1, Appli  
37 1132 45.8 9719 4 US-09-700-304-1 Sequence 1, Appli  
38 1128.8 45.7 9050 2 US-09-184-418C-7 Sequence 7, Appli  
39 1125.6 45.6 7399 2 US-08-418-848A-9 Sequence 9, Appli  
40 1125.6 45.6 9709 2 US-08-188-583-5 Sequence 5, Appli  
41 1125.6 45.6 9709 3 US-08-388-353-1 Sequence 1, Appli  
42 1125.6 45.6 9709 3 US-08-488-551B-1 Sequence 1, Appli  
43 1125.6 45.6 9709 3 US-09-309-572-15 Sequence 15, Appli  
44 1125.6 45.6 9709 4 US-09-718-096-15 Sequence 15, Appli  
45 1125.6 45.6 12479 4 US-09-318-138-13 Sequence 13, Appli

## ALIGNMENTS

## RESULT 1

US-09-475-515-84  
; Sequence 84, Application US/09475515A  
; Patent NO. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 84  
; LENGTH: 2312  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(-).protmod.RTopt(+)  
US-09-475-515-84

Query Match 83.1%; Score 2052; DB 4; Length 2312;  
Best Local Similarity 93.6%; Pred. No. 0;  
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
Qy 170 GCGCAAGGAGGGCCACGAGTGAAGGACTGCAGCGCGCGAGCCAGCTTCTTCGCG 229  
Db 1 CGGCGCGGAGGACCAACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60  
Qy 230 AGGACCTTGGCTTTCCCCAGGGCAAGCCCGCGAGTTTCCCCAGCGAGCAGAACCGCGCA 289  
Db 61 AGGACCTTGGCTTTCTTCGAGGCAAGCCCGCGAGTTTCAGCAGCGAGCAGACCCGCGCA 120  
Qy 290 ACAGCCCCACAGCGCGAGCTGCAGTGGCGCGG-----ACAACCCCGAGCGAGG 343  
Db 121 ACAGCCCCACCGCGCGAGCTGCAGTGTGGGCGCGGAGAACAAAGCCTTGAGCGAGG 180  
Qy 344 CCGCGCGCGAGCGCGAGGCGACCCCTG-----AACTTCCCCCAGATCACCTCTGGGCGAGC 397  
Db 181 CCGCGCGCGAGCGCGGCGGACCCCTGTAGCTTCACTTCCCCCAGATCACCTCTGTGGCAGC 240

QY	398	GCC	CCT	TGG	TGAG	CAT	CA	AG	TGG	GGG	CCG	CA	GAT	CA	AG	AGG	GGC	CT	TG	TG	GA	CAC	CGG	C	457	
DB	241	GCC	CCT	TGG	TAC	CA	CT	CAG	GA	T	CGG	CGG	CCAG	CT	CA	AG	AGG	GGG	CG	T	CT	CG	A	CGG	C	300
QY	458	CCG	ACG	CAC	CGT	GCT	GAG	GAG	AT	GAC	CT	GCC	CGG	CA	AGT	GGA	AGC	CC	CA	A	GAT	GAT	CG		517	
DB	301	CCG	ACG	CAC	CGT	GCT	GAG	GAG	AT	GAA	CT	TG	CCG	CGG	CA	AGT	GGA	AGC	CC	CA	A	GAT	CG		360	
QY	518	GCG	GAT	CCG	GGC	CTT	CAT	CA	AGT	TGG	CGC	AGT	AC	GAC	CA	GAT	CCT	T	GAT	CG	A	GAT	CT	CGC	577	
DB	361	GCG	GAT	CCG	GGC	CTT	CAT	CA	AGT	TGG	CGC	AGT	AC	GAC	CA	GAT	CCT	T	GAT	CG	A	GAT	CT	CGC	420	
QY	578	GCA	AGA	AGC	CCAT	TGG	CA	CCG	TG	TGAT	TGG	CCCC	CA	CCCC	CGT	GAA	CAT	CAT	TGG	CG	CGCA			637		
DB	421	GCC	ACA	AGC	CCAT	TGG	CA	CCG	TG	TGAT	TGG	CCCC	CA	CCCC	CGT	GAA	CAT	CAT	TGG	CG	CGCA			480		
QY	638	ACAT	GCT	GAC	CCAG	CTGG	GCT	GAC	CCCT	GAA	CTT	CC	CCAT	CAG	CCCC	CCAT	CG	A	GAC	CGT	GC			697		
DB	481	ACCT	GCT	GAC	CCAG	ATCG	GCT	GCA	CCCT	GAA	CTT	CC	CCAT	CAG	CCCC	CCAT	CG	A	GAC	CGT	GC			540		
QY	698	CCG	TGA	AGT	CTA	AG	CCCG	GCAT	GCA	CGC	CCC	CA	AGGT	GAA	GCA	GCT	GAC	CGT	CCCT	CA	CCG	AGG		757		
DB	541	CCG	TGA	AGT	CTA	AG	CCCG	GCAT	GCA	CGC	CCC	CA	AGGT	GAA	GCA	GCT	GAC	CGT	CCCT	CA	CCG	AGG		600		
QY	758	AGA	AGAT	CA	AGC	CCCT	GAC	CGCC	ATCT	CG	AGG	AGAT	TG	GAA	AGG	AGG	CA	A	GAT	CA	CCA			817		
DB	601	AGN	AGAT	CA	AGC	CCCT	TGT	GAG	ATCT	GCA	CCG	AGAT	TG	GAA	AGG	AGG	CA	A	GAT	CA	GCA			660		
QY	818	AGAT	CGG	CCCC	GAG	AA	CCCC	CTA	CA	AC	CCCC	CGT	TG	TG	CCAT	CA	AGA	AGA	AGG	CA	GCA			877		
DB	661	AGAT	CGG	CCCC	GAG	AA	CCCC	CTA	CA	AC	CCCC	CGT	TG	TG	CCAT	CA	AGA	AGA	AGG	CA	GCA			720		
QY	878	CCA	AGT	G	CGCA	AG	CTGT	G	ACT	TT	CG	CGAG	CT	GAA	CA	AGC	G	CAC	CC	AG	ACT	TT	CTGG	937		
DB	721	CCA	AGT	G	CGCA	AG	CTGT	G	ACT	TT	CG	CGAG	CT	GAA	CA	AGC	G	CAC	CC	AG	ACT	TT	CTGG	780		
QY	938	AGT	G	CAG	CTGG	GCAT	CCCC	CA	CCCC	CGC	CGC	CTT	GAA	AGA	AGA	AGG	AGG	CGT	G	AC	CGT	GC		997		
DB	781	AGT	G	CAG	CTGG	GCAT	CCCC	CA	CCCC	CGC	CGC	CTT	GAA	AGA	AGA	AGG	AGG	CGT	G	AC	CGT	GC		840		
QY	998	TGG	AG	CTGGG	CGA	CG	CTT	AC	AG	CTG	CGC	CTT	GGA	AGG	AGG	ACT	T	CGG	CA	AGT	CA	CCG		1057		
DB	841	TGG	AG	CTGGG	CGA	CG	CTT	AC	AG	CTG	CGC	CTT	GGA	AGG	AGG	ACT	T	CGG	CA	AGT	CA	CCG		900		
QY	1058	CCTT	CAC	CAT	CCC	CAG	CAT	CA	A	CA	CAG	AC	CCC	CGC	GCAT	CCG	CT	TAC	AGT	CA	CA	CGT	GC	1117		
DB	901	CCTT	CAC	CAT	CCC	CAG	CAT	CA	A	CA	CAG	AC	CCC	CGC	GCAT	CCG	CT	TAC	AGT	CA	CA	CGT	GC	960		
QY	1118	TG	CCCC	CAG	GGT	TG	GA	GGC	AG	CCCC	CA	G	CAT	TTT	CC	AG	G	CAG	CAT	G	CA	A	GAT	CT	CGG	1177
DB	961	TG	CCCC	CAG	GGT	TG	GA	GGC	AG	CCCC	CA	G	CAT	TTT	CC	AG	G	CAG	CAT	G	CA	A	GAT	CT	CGG	1020
QY	1178	AG	CCCTT	CCG	CGG	CGCA	AA	CCCC	CAG	AT	CTGT	GAT	CTA	CC	AGT	A	CAT	TGGA	GCA	CT	GT	PAC		1237		
DB	1021	AG	CCCTT	CCG	CAAG	CAG	AA	CCCC	CAC	AT	CTGT	GAT	CTA	CC	AGT	A	CAT	TGGA	GCA	CT	GT	PAC		1080		
QY	1238	TGG	G	CAG	CGA	CT	GAG	AT	CGG	C	GAG	CAC	CG	CGC	CA	AGAT	CG	GAG	AGT	CG	GCA	AGC	CAC	1297		
DB	1081	TGG	G	CAG	CGA	CT	GAG	AT	CGG	C	GAG	CAC	CGC	CA	AGAT	CG	GAG	AGT	CG	GCA	AGC	CAC	1140			
QY	1298	TGCT	G	CGCT	TGG	GGCTT	CAC	CA	CCCC	CGCA	AA	GA	AGC	AC	CA	GA	AGG	AGG	CCCC	CCCC	TT	CTCT	GT	1357		
DB	1141	TGCT	G	CGCT	TGG	GGCTT	CAC	CA	CCCC	CGCA	AA	GA	AGC	AC	CA	GA	AGG	AGG	CCCC	CCCC	TT	CTCT	GT	1200		
QY	1358	GGAT	TGG	GGT	CTA	CG	AGT	GCA	CCCC	CGCA	AA	GTGGA	AC	CGT	TG	AG	CC	CA	CT	CG	AG	CT	CGG	AG	1417	
DB	1201	GGAT	TGG	GGT	CTA	CG	AGT	GCA	CCCC	CGCA	AA	GTGGA	AC	CGT	TG	AG	CC	CA	CT	CG	AG	CT	CGG	AG	1260	
QY	1418	AGG	AG	AGCT	TG	CAC	CGT	GAA	CC	CA	CT	CCAG	AG	CT	G	TGG	G	CA	AG	CT	G	CA	CT	TGG	CG	1477
DB	1261	AGG	AG	AGCT	TG	CAC	CGT	GAA	CC	CA	CT	CCAG	AG	CT	G	TGG	G	CA	AG	CT	G	CA	CT	TGG	CG	1320
QY	1478	AGAT	CTA	CCCC	CGG	CAT	CA	AGT	TGG	CGG	CCAG	CT	TG	TG	CA	AG	AGT	CT	G	CT	CG	GG	GG	CG	1537	

RESULT 2  
US-09-477  
; Sequen  
; Patent  
; GENERA

	: APPLICANT: BARNETT, Susan	
	: APPLICANT: ZUR MEGEDE, Jan	
	: APPLICANT: SRIVASTAVA, Indresh	
	: APPLICANT: LIAN, Ying	
	: APPLICANT: HARTOG, Karin	
	: APPLICANT: LIU, Hong	
	: APPLICANT: GREER, Catherine	
	: APPLICANT: SELBY, Mark	
	: APPLICANT: WALKER, Christopher	
	: TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION	
	: TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES	
	: FILE REFERENCE: 1621.002	
	: CURRENT APPLICATION NUMBER: US/09/475,515A	
	: CURRENT FILING DATE: 1999-12-30	
	: NUMBER OF SEQ ID NOS: 90	
	: SOFTWARE: PatentIn Ver. 2.0	
	: SEQ ID NO 82	
	: LENGTH: 2306	
	: TYPE: DNA	
	: ORGANISM: Artificial Sequence	
	: FEATURE:	
	: OTHER INFORMATION: Description of Artificial Sequence:	
	: OTHER INFORMATION: FS(-).protmod.RTopt.YM	
	US-09-475-515-82	
	Query Match 82.0%; Score 2025.2; DB 4; Length 2306;	
	Best Local Similarity 93.3%; Pred. No. 0;	
	Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;	
Qy	170 GCGCCAAGGCGCCACCAAGATGTAAGACTGCACCAGGCGCCAGGCCAACTTCTTCGGC 229	
Dd	1 GCGCCGCGGAAGACACCAAATAAAGATTGCACTGAGAGACAGGCTAATTTCTTCGGC 60	
Qy	230 AGGACCTGGGCTTCCCACAGGCGAAGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCA 289	
Dd	61 AGGACCTGGGCTTCTGCGAGGCGAAGCCCGCGAGTTGAGCAGCGAGCACCCGCGCA 120	
Qy	290 ACAGCCCAACAGCCCGAGCTGCGAGTGCGCGCG-------ACAACCCCGCAGCGAGG 343	
Dd	121 ACAGCCCAACAGCCCGAGCTGCGAGTGCGCGCGCGAGAACACAGCCTGAGCGAGG 180	
Qy	344 CCGCGCGAGCGCGCAGGCGCACCTG-----AATTCCCCCAGATCACCTGTGGCAGC 397	
Dd	181 CCGCGCGAGCGCGCAGGCGCACCTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC 240	
Qy	398 GCCCCTGTGTAGCATCAAGGTGGGCGCGCAGATCAAGAGSGCCCTGTGGACACCGGCG 457	
Dd	241 GCCCCTGTGTAGCATCAGATCGGCGCGCAGCTCAAGAGSGCGCTGTGCACACCGGCG 300	
Qy	458 CCAGCAGCACCGTGTGGAGGAGATGAGCTCCCGCGCAAGTGGAAAGCCCAAGATGATCG 517	
Dd	301 CCAGCAGCACCGTGTGGAGGAGATGAACCTCCCGCGCAAGTGGAAAGCCCAAGATGATCG 360	
Qy	518 CGCGCATCGCGCGCTTCATCAAGGTGCGCGCAGTAGACACAGATCTCTGATCGAGATCTGCG 577	
Dd	361 CGCGCATCGCGCGCTTCATCAAGGTGCGCGCAGTAGACACAGATCTCTGATCGAGATCTGCG 420	
Qy	578 GCAGAGGCGCATCGGCGCGCTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637	
Dd	421 GCACAGGCGCATCGGCGCGCTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 480	
Qy	638 ACATGTGTGACCCAGCTGGGCTGACCCCTGAACTTCCCCCATCAGCCCCCATCGAGACCGTGC 697	
Dd	481 ACCTGTGTGACCCAGATCGGCTGACCCCTGAACTTCCCCCATCAGCCCCCATCGAGACCGTGC 540	
Qy	698 CCGTGAAGTGAAGCCCGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGG 757	
Dd	541 CCGTGAAGTGAAGCCCGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGG 600	
Qy	758 AGAAGATCAAGGCCCTTGAACCGGATCTGCGAGAGATGAGAGAGGCGGCGAGATCACCA 817	
Dd	601 AGAAGATCAAGGCCCTTGTGGAGATCTGCACCGGATGAGAGAGGCGGCGAGATCACCA 660	



```

Db 1735 CCTGGATCCCGAGTGGAGTCTGTGAACACCCCCCTGTGTGAAGCTGTGTACAGC 1794
QY 1958 TGAGAAAGAGCCCATCATCGGCGCCGAGACCTTCTAGTGGAGCGGCGGCCAACCGCG 2017
Db 1795 TGAGAAAGAGCCCATCATCGTGGCGCGGAGACCTTCTACGTGGACGGCGCGCCAACCGCG 1854
QY 2018 AGACCAAGATCCGCAAGCGCGGTACGTGACCGACCGGCGCGGAGAGATCGTGAGCC 2077
Db 1855 AGACCAAGCTTGGCAAGAGCGCGGTCTAGTGACCCAGCCGGGCGCGGAGAGGTGGTGAACA 1914
QY 2078 TGACCCAGACACCAACCAAGAGACCGAGCTGACCGAGCCGAGCTGCCAGCTGGCCCTTCAGGAGCA 2137
Db 1915 TGCCCGACACCAACCAAGAGACCGAGCTGACCGAGCCGAGCTGCCAGCTGGCCCTTCAGGAGCA 1974
QY 2138 GCGGACGAGAGGTGAACATCTGTGACCGAGCAGCCAGTACGCGCTTGGGAGATCATCCAGGCC 2197
Db 1975 GCGGCTTGGAGGTGAACATCTGTGACCGAGCAGCCAGTACGCGCTTGGGAGATCATCCAGGCC 2034
QY 2198 AGCCGACAGAGCGAGAGCGAGCTGTGTGAACCGAGATCATCGAGCAGCTGATCAAGAAGG 2257
Db 2035 AGCCGACAGAGCGAGAGCGAGCTGTGTGAACCGAGATCATCGAGCAGCTGATCAAGAAGG 2094
QY 2258 AGAAGGTGTACCTGAGCTGGGTGCCCGCCCAAGCGGCATCGCGCGCAACGAGCAGATCG 2317
Db 2095 AGAAGGTGTACCTGGCTGGGTGCCCGCCCAAGCGGCATCGCGCGCAACGAGCAGGTGG 2154
QY 2318 ACAAGCTGTGAGCAAGGCGATCCGCAAGTGTGTCTTGGAGCGGATCGATGGCGGCA 2377
Db 2155 ACAAGCTGTGAGCGCGGCATCCGAAGGTGTGTCTTGAACGGCAGTGTGGCGGCA 2214
QY 2378 TCGTGATCTACCAAGTACAGGAGCTGTGACGCTGTGGCGAGCGCGGCGCTTAGGATCGATT 2437
Db 2215 TCGTGATCTACCAAGTACAGGAGCTGTGACGCTGTGGCGAGCGCGGCGCTTAGGATCGATT 2274
QY 2438 AAAAGCTTCCCGGGGTAGCACCGGTGAATTC 2469
Db 2275 AAAAGCTTCCCGGGGTAGCACCGGTGAATTC 2306

```

```

RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEDEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; APPLICANT: WALKER, Christopher
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475, 515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.RTopt.YMMW
US-09-475-515-83

```

Query Match 80.9%; Score 1998.4; DB 4; Length 2300;  
 Best Local Similarity 92.9%; Pred. No. 0;  
 Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

```

QY 170 GCGGCAAGGAGGCGCCACAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGCG 229
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGACTGAGAGACAGGCTAATTTCTTCGCG 60
QY 230 AGGACCTGCGCTTCCCGCAGGGCAAGGCGCGAGTTCCCGCAGCAGAGAAACCGGCGCA 289
Db 61 AGGACCTGCGCTTCCCGCAGGGCAAGGCGCGAGTTCCAGCAGCAGAGACCGCGCGCA 120
QY 290 AAGCCCCACAGCCCGAGACTGCAAGTGGCGGGG-----ACAACCCCGCAGCGAG 343
Db 121 ACAGCCCCACCGCCCGAGCTGCAAGTGTGGGGCGGCGAGAACAAACAGCTGAGCGAG 180
QY 344 CCGGCGCCGAGCGCCAGGGCACCCCTG-----AACTTCCCGCAGATCACCTCTGTGGCAG 397
Db 181 CCGGCGCCGAGCGCCAGGGCACCGTGAGCTTCACTTCCCGCAGATCACCTCTGTGGCAG 240
QY 398 GCGCCCTGTGTAGCATCAAGGTGGCGCCAGATCAAGGAGGCCCTGTCTGGAACCGCGCG 457
Db 241 GCGCCCTGTGTAGCATCAAGGTGGCGCCAGCTCAAGGAGGCCCTGTCTGGAACCGCGCG 300
QY 458 CCGACGACACCGCTGTGGAGGAGATGAGCTGCGCGGCAAGTGGAGGCCCAAGATGATCG 517
Db 301 CCGACGACACCGCTGTGGAGGAGATGAGCTGCGCGGCAAGTGGAGGCCCAAGATGATCG 360
QY 518 GCGGATCGCGCGCTTTCATCAAGGTGGCGCAGTACGACGAGATCTCTGATCGAGATCTGCG 577
Db 361 GCGGATCGCGCGCTTTCATCAAGGTGGCGCAGTACGACGAGATCTCTGATCGAGATCTGCG 420
QY 578 GCAAGAGGCCCATCGGACACCGTGTGATCGGCGCCACCGCGTGAACATCATCGGCGCGCA 637
Db 421 GCACAAAGGCCCATCGGACACCGTGTGATCGGCGCCACCGCGTGAACATCATCGGCGCGCA 480
QY 638 ACATGCTGACCGAGCTGGGCTGCACTTCCCGATCAGCCCGATCAGCCCGATCAGACCGTGC 697
Db 481 ACCTGCTGACCGAGATCGGCTGCACTTCCCGATCAGCCCGATCAGACCGGTCG 540
QY 698 CCGTGAAGCTGAAGCGCGCATGCGCGCCCAAGGTGAAGCAGTGGCGCCCTCAGCGAGG 757
Db 541 CCGTGAAGCTGAAGCGCGGATGCGCGCCCAAGGTGAAGCAGTGGCGCCCTCAGCGAGG 600
QY 758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAAGGAGGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCCCTGTTGGAGATCTGCACCGAGATGGAGAAAGGAGGCAAGATCAGCA 660
QY 818 AGATCGGCGCGAGAAACCCCTTCAACACACCGCGTGTTCGCCATCAGAGAAAGAGCAGCA 877
Db 661 AGATCGGCGCGAGAAACCCCTTCAACACACCGCGTGTTCGCCATCAGAGAAAGAGCAGCA 720
QY 878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGG 937
Db 721 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGGG 780
QY 938 AGGTGAGCTGGGCTATCCCGCCACCGCGCGCTGAAGAAAGAGAGCGTCAACCGTGC 997
Db 781 AGGTGAGCTGGGCTATCCCGCCACCGCGCGCTGAAGAAAGAGAGCGTCAACCGTGC 840
QY 998 TGGACGTGGGCGAGCGCTTCTTTCAGCGTGGCGCGCGCTTTCAGCGAGGACTTTCGCAAGTACACCG 1057
Db 841 TGGACGTGGGCGAGCGCTTCTTTCAGCGTGGCGCGCGCTTTCAGCGAGGACTTTCGCAAGTACACCG 900
QY 1058 CTTTACCATCCCCAGCATCAACAAAGCAGACCCCGCGCATCTCCGCTACCAAGTACAAAGTGC 1117
Db 901 CTTTACCATCCCCAGCATCAACAAAGCAGACCCCGCGCATCTCCGCTACCAAGTACAAAGTGC 960
QY 1118 TGCCCCAGGGCTGGAAGGCGAGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTGTG 1177
Db 961 TGCCCCAGGGCTGGAAGGCGAGCCCGCGCATCTTCAGAGCAGCATGACCAAGATCTGTG 1020
QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATGTGTATCTTACAGTACATGAGAGCTGTACG 1237
Db 1021 AGCCCTTCCGCGCGCGCAAGAACCCCGAGATCGTGTATCTACCA-----GGCCCCCTGTACG 1074

```

1238 TGGGACGCGA CTTGGAGATCGGCAGACACCGGCCAGATCGAGAGCTGCGCAGCACC 1297  
Db |  
1075 TGGGACGCGA CTTGGAGATCGGCAGACACCGGCCAGATCGAGAGCTGCGCAGCACC 1134  
Qy |  
1298 TGCTGCGCTGGGGCTTACCAACCCCGGCAAGAAGCACCAGAAGAGCCGCCCTTCCTGT 1357  
Db |  
1135 TGCTGCGCTGGGGCTTACCAACCCCGGCAAGAAGCACCAGAAGAGCCGCCCTTCCTGC 1194  
Qy |  
1358 GGATGGGCTACGAGTGCACCCCGGCAAGAAGTGGACCGTGCAGCCCATCGAGTGCGCCGAGA 1417  
Db |  
1195 CCAT-----CGAGCTGACCCCGGCAAGAAGTGGACCGTGCAGCCCATCGAGTGCGCCGAGA 1248  
Qy |  
1418 AGGAGAGCTGCACCGTGAACACATCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCCAGCC 1477  
Db |  
1249 AGGAGAGCTGCACCGTGAACACATCCAGAGAGCTGGTGGGCAAGCTGAACCTGGGCCAGCC 1308  
Qy |  
1478 AGATCTACCCCGGCATCAAGCTGGGCCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCC 1537  
Db |  
1309 AGATCTACCCCGGCATCAAGCTGAAGCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCC 1368  
Qy |  
1538 TGACCGACATCGTGCCTGTGACCGAGGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGA 1597  
Db |  
1369 TGACCGAGGTGATCCCTGTGACCGAGGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGA 1428  
Qy |  
1598 TCCTGGCGGAGCCGTGTCACCGGTGTACTACGACCCAGCAGCAGCAGCTGTGGCCGAGA 1657  
Db |  
1429 TCCTGAAGGAGCCGTGTCACCGGTGTACTACGACCCAGCAGCAGCAGCTGTGGCCGAGA 1488  
Qy |  
1658 TCCAGAGAGCGGCCACGACAGCTGGACCTACACAGATCTACGAGAGCCCTTCAAGAAC 1717  
Db |  
1489 TCCAGAGAGCGGCCACGACAGCTGGACCTACACAGATCTACGAGAGCCCTTCAAGAAC 1548  
Qy |  
1718 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCAACAGCAGTGAAGCAGCTGA 1777  
Db |  
1549 TGAAGACCGGCAAGTACGCCAAGATGCGCACCGGCCACACCAACAGCAGTGAAGCAGCTGA 1608  
Qy |  
1778 CCGAGGCGGTGCAGAAGATCCCATGAGAGAGATCTGTATCTGGGGCAAGACCCCAAGT 1837  
Db |  
1609 CCGAGGCGGTGCAGAAGATCCCATGAGAGAGATCTGTATCTGGGGCAAGATCCCAAGT 1668  
Qy |  
1838 TCCGCTTCCCATCCAGAGAGACCTGGGAGACCTGGTGACCGACTACTTGGCAGGCA 1897  
Db |  
1669 TCAAGCTGCCCATCCAGAGAGACCTGGGAGGCTGTGTGATGTAGTACTTGGCAGGCA 1728  
Qy |  
1898 CCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACCGAGC 1957  
Db |  
1729 CCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACCGAGC 1788  
Qy |  
1958 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTAGTGGACGGCGCCGCCAACCGCG 2017  
Db |  
1789 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTAGTGGACGGCGCCGCCAACCGCG 1848  
Qy |  
2018 AGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCCGGGCGGCGAGAGATCTGAGCC 2077  
Db |  
1849 AGACCAAGCTGGGCAAGCGCGCTACGTGACCGAGCCGGGCGGCGAGAGTGTGAGCA 1908  
Qy |  
2078 TGACCGGAGACCAACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGGACA 2137  
Db |  
1909 TCGCCGACACCAACCAACAGAGACCGAGCTGCAGGCGCATCCAGTGGCCCTGCAGGACA 1968  
Qy |  
2138 GCGCAGCGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCGATCATCAGGCCC 2197  
Db |  
1969 GCGGCTGGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCGATCATCAGGCCC 2028  
Qy |  
2198 AGCCCGACAGAGCGAGCGAGCTGTGTGAACACAGATCATCGACAGCTGATCAAGAGG 2257  
Db |  
2029 AGCCCGACAGAGCGAGCGAGCTGTGTGAACACAGATCATCGACAGCTGATCAAGAGG 2088  
Qy |  
2258 AGAAGGTGTACTGAGCTGGGTGCGCGCCCAAGAGGCGATCGCGGCAACAGCAGATCG 2317  
Db |  
2089 AGAAGGTGTACTGAGCTGGGTGCGCGCCCAAGAGGCGATCGCGGCAACAGCAGAGGTGG 2148  
Qy |  
2318 ACAAGCTGGTGAACAGGGCATCCGCAAGGTGCTGTTCTTGGCAAGCGGCATCGATGGCGGCA 2377

Db |  
2149 ACAAGCTGGTGGGCGCGGCATCCGCAAGGTGCTTCTTGAACGGCATCGATGGCGCA 2208  
Qy |  
2378 TCCTGATCTACCACTACATGACGACCTGTACGTGGGCGAGCGCGCCCTTAGGATCGATT 2437  
Db |  
2209 TCCTGATCTACCACTACATGACGACCTGTACGTGGGCGAGCGCGCCCTTAGGATCGATT 2268  
Qy |  
2438 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2469  
Db |  
2269 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2300

## RESULT 4

US-09-475-515-6  
; Sequence 6, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MESEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 4319  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-Gag-polymerase  
US-09-475-515-6

Query Match 79.8%; Score 1969.4; DB 4; Length 4319;

Best Local Similarity 90.1%; Pred. No. 0;

Matches 2158; Conservative 0; Mismatches 221; Indels 16; Gaps 4;

Qy 14 TGGCCGAGGCCATCAGCCAGCCAGCCAGC---GCCAATCCTCTGATGAGCGCAGCAACT 70

Db 1100 TGGCCGAGGCCATCAGCCAGCCAGCCAGC---GCCAATCCTCTGATGAGCGCAGCAACT 1159

Qy 71 TCAAGGGCCCCCAAGCGCATCATCAAGTCTTCAACTGCGCAAGAGGGCCACATCGCCC 130

Db 1160 TCCGCAACCGAGCGGAAGACCGTCAAGTCTTCACTGCGCAAGAGGGCCACACCGCA 1219

Qy 131 GCAATCGCGCGCCCCCGCAAGAGGCTCTGGAAGTGGCGAAGAGGGCCACCAAGA 190

Db 1220 GGAATCGCGCGCCCCCGCAAGAGGCTCTGCGCTGCGCGCGCGAAGAGCACCAAA 1279

Qy 191 TGAAGGACTGACCGAGCGCCAGCCCACTTCTCCGCGAGGACTTGGCCCTTCCCCCAGG 250

Db 1280 TGAAGAGTTGACTGAGAGACAGGCTAA-TTTTATGGGAAGATCTGGCCCTTCTTACAG 1338

Qy 251 GCAAGGCGCGGAGTCTCCCGAGCGAGAGAACCGCGCAACAGCCCAACAGCGCGCAGC 310

Db 1339 GGAAGGCCAGGGAATTTCTTCAGAGCAGACAGCCACAGCCCAACAGAGAGAGAGC 1398

Qy 311 TGCAGGTGCGCGCG-----ACAAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCA 364

Db 1399 TTCAGGTGTTGGGAGGAGAGAAACAACTCCCTCTCAGAAGCAGAGGCGCGGATAGACAAAGAA 1458

Qy 365 -----CCCTGAACCTTCCCGAGATCACTCTGTCGAGCGCGCCCTTGTGAGCATCAGG 418

Db 1459 CTGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAACGAGCCCTCTGTACAGTAAGGA 1518

QY 419 TGGCGGCGCAGATCAAGAGGCGCTGCTGTGACACCGCGCGCGACACCGCTGCTGAGG 478  
Db 1519 TCGCGCGCCAGCTCAAGAGGCGCTGCTGTGACACCGCGCGCGACACCGCTGCTGAGG 1578  
QY 479 AGATGAGCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCA 538  
Db 1579 AGATGAACCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCA 1638  
QY 539 AGTGGCGCAGTAGGACAGATCTGTATCGAGATCTTGGCGCAAGAGGCCATCGGCACCG 598  
Db 1639 AGTGGCGCAGTAGGACAGATCTGTATCGAGATCTTGGCGCAAGAGGCCATCGGCACCG 1698  
QY 599 TGTGTATCGGCGCCACCGCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGGCT 658  
Db 1699 TGTGTGTGGCGCCACCGCGTGAACATCATCGGCGCAACATGCTGACCCAGATCGGCT 1758  
QY 659 GCACCTGTGAATTTCCCATCAGCCCATCGAGCCGTCGCGGTGAAGCTGAAGCCCGGCA 718  
Db 1759 GCACCTGTGAATTTCCCATCAGCCCATCGAGCCGTCGCGGTGAAGCTGAAGCCCGGCA 1818  
QY 719 TGAACGCGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAAGATCAAGGCCCTGACCG 778  
Db 1819 TGAACGCGCCCAAGGTGAAGCAGTGGCCCTGTGACCGAGGAGAAGATCAAGGCCCTGAGTG 1878  
QY 779 CCATCTGCGAGGAGATGGAAAGGAGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCT 838  
Db 1879 AGATCTGACCGAGATGGAAAGGAGGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCT 1938  
QY 839 ACAACACCCCGTGTTCGCGCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTG 898  
Db 1939 ACAACACCCCGTGTTCGCGCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTG 1998  
QY 899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTTCTGGGAGGTGAGCTGGGGCATCCCC 958  
Db 1999 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTTCTGGGAGGTGAGCTGGGGCATCCCC 2058  
QY 959 ACCCGCGCGCTGAGAAAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACT 1018  
Db 2059 ACCCGCGCGCTGAGAAAGAGAGCGTGACCGTGTGGAGCTGGGCGAGCGCTACT 2118  
QY 1019 TCAGCGTGCCTTCGAGCGAGGACTTCGCGAAGTACACCGCTTTACCCATCCCCAGCATCA 1078  
Db 2119 TCAGCGTGCCTTCGAGCGAGGACTTCGCGAAGTACACCGCTTTACCCATCCCCAGCATCA 2178  
QY 1079 ACAACGAGACCCCGGATCCGCTACAGTACAAAGTGTGCTGCTGCGCGAGGCTGAAGGGCA 1138  
Db 2179 ACAACGAGACCCCGGATCCGCTACAGTACAAAGTGTGCTGCTGCGCGAGGCTGAAGGGCA 2238  
QY 1139 GCCCGAGCATTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGCAACC 1198  
Db 2239 GCCCGAGCATTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCGCGCAACC 2298  
QY 1199 CCGAGATCTGTGATCTACAGTACATGAGACGACCTGTACGTGGGCGAGGACTGGAGATCG 1258  
Db 2299 CCGACATCTGTGATCTACAGTACATGAGACGACCTGTACGTGGGCGAGGACTGGAGATCG 2358  
QY 1259 GCAGACACCGCGCAAGATCGAGGAGCTGCGGAGACCTGTGCTGCGCTGGGGCTTACCA 1318  
Db 2359 GCAGACACCGCGCAAGATCGAGGAGCTGCGGAGACCTGTGCTGCGCTGGGGCTTACCA 2418  
QY 1319 CCCCAGCAGAGAGCACCAAGAGGCGCCCTCTCTGTGGATGGGCTAGAGCTGACCC 1378  
Db 2419 CCCCAGCAGAGAGCACCAAGAGGCGCCCTCTCTGTGGATGGGCTAGAGCTGACCC 2478  
QY 1379 CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGGAGAGCTGGACCGTGAACG 1438  
Db 2479 CCGACAAGTGGACCGTGCAGCCCATCATGTGCGCGAGAGGAGACAGCTGGACCGTGAACG 2538  
QY 1439 ACATCCAGAGCTGTGGGCAAGCTGAACTGTGGGCGAGCAGATCTACCCCGGATCAAG 1498  
Db 2539 ACATCCAGAGCTGTGGGCAAGCTGAACTGTGGGCGAGCAGATCTACCCCGGATCAAG 2598

QY 1499 TGGCCAGCTGTGAAAGCTGCTGCGCGGCGCCAAAGGCCCTTGACCGGACATCGTGCCTCTGA 1558  
Db 2599 TGAAGCAGCTGTGCAAGCTGCTGCGCGGCGCACCAAGGCCCTTGACCGAGGTGATCCCCCTGA 2658  
QY 1559 CCAGAGAGCCGAGCTGAGAGCTGGCGCGAGAACCCGCGAGATCTCTTGGCGGAGACCCGCTGACG 1618  
Db 2659 CCAGAGAGCCGAGCTGAGAGCTGGCGCGAGAACCCGCGAGATCTCTTGAAGGAGCCCGCTGACG 2718  
QY 1619 GCGTGTACTACGACCCCGACGACAGGACCTGTGTGCGCGAGATCTCCAGAGAGCAGGGCCACGACC 1678  
Db 2719 AGTGTGTACTACGACCCCGACGACAGGACCTGTGTGCGCGAGATCTCCAGAGAGCAGGGCC 2778  
QY 1679 AGTGTGACCTTACAGAGATCTTACGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCA 1738  
Db 2779 AGTGTGACCTTACAGAGATCTTACGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCA 2838  
QY 1739 AGATGGCGACCGGCCACACCAACGAGCTGAGAGCAGCTGACCGAGGCGCTGCGAGAGATCG 1798  
Db 2839 GCATGCGCGCGCGCCACACCAACGAGCTGAGAGCAGCTGACCGAGGCGCTGCGAGAGGTGA 2898  
QY 1799 CCATGAGAGAGCATCGTGTGATCTGGGCGAAGACCCCAAGTTTCCGCTGCCCATCCAGAAAG 1858  
Db 2899 GCATGAGAGAGCATCGTGTGATCTGGGCGAAGATCCCAAGTTTCAAGCTGCCCATCCAGAAAG 2958  
QY 1859 AGACCTGGGAGAGCATCGTGTGATCTGGGCGAAGACCCCAAGTTTCCGCTGCCCATCCAGAAAG 1918  
Db 2959 AGACCTGGGAGAGCATCGTGTGATCTGGGCGAAGATCCCAAGTTTCAAGCTGCCCATCCAGAAAG 3018  
QY 1919 TCGTGAACACCCCGCTGCTGTGAAAGCTGTGTGATCCAGCTGTGAGAGAGGAGCCCATCATCG 1978  
Db 3019 TCGTGAACACCCCGCTGCTGTGAAAGCTGTGTGATCCAGCTGTGAGAGAGGAGCCCATCATCG 3078  
QY 1979 GCGCGAGAGCTTCTACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCG 2038  
Db 3079 GCGCGAGAGCTTCTACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCG 3138  
QY 2039 GCTACGTGACCGACCGGCGCGCGCAGAGAAGATCGTGTGAGCTGTGAGAGACCAACACAG 2098  
Db 3139 GCTACGTGACCGACCGGCGCGCGCAGAGAAGTGTGTGAGCATCGCGGACACCAACACAG 3198  
QY 2099 AGACCGAGCTGAGGCGCATCTCAGCTGGCCCTTGAGGACAGCGGCGAGGAGTGAACATCG 2158  
Db 3199 AGACCGAGCTGAGGCGCATCTCAGCTGGCCCTTGAGGACAGCGGCGCTGGAGGTGAACATCG 3258  
QY 2159 TGAACGACAGCAGTACGCGCTGGGCGATCATCCAGGCCAGCGCCGACCAAGAGCGAGAGCG 2218  
Db 3259 TGAACGACAGCAGTACGCGCTGGGCGATCATCCAGGCCAGCGCCGACCAAGAGCGAGAGCG 3318  
QY 2219 AGCTGTGTAAACCAAGATCATCGAGCAGCTGATCAAGAGAGGAGAGGTGTACTGTAGCTGGG 2278  
Db 3319 AGCTGTGTAGCCAGATCATCGAGCAGCTGATCAAGAGAGGAGAGGTGTACTGTGGCTGGG 3378  
QY 2279 TGGCGCGCCCAAGAGGCGATCGGCGGCAACGAGCAGATCGCAAGCTGGTGAAGAGGGCA 2338  
Db 3379 TGGCGCGCCCAAGAGGCGATCGGCGGCAACGAGCAGGTGGAACAAGCTGGTGAAGCGCGCA 3438  
QY 2339 TCCGCAAGGTGCTGTCTTCCCTGGAGCGGATCGATGCGCGCATCTGTGATCTACAGTA 2393  
Db 3439 TCCGCAAGGTGCTGTCTTCCCTGGAGCGGATCGCAAGGCCCGAGGAGGACACAGAA 3493

## RESULT 5

US-09-475-515-80  
; Sequence 80, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine

APPLICANT: SELBY, Mark  
APPLICANT: WALKER, Christopher  
TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
FILE REFERENCE: 1621.002  
CURRENT APPLICATION NUMBER: US/09/475.515A  
CURRENT FILING DATE: 1999-12-30  
NUMBER OF SEQ ID NOS: 90  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 80  
LENGTH: 2305  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
OTHER INFORMATION: PS(+).prolnact.RTopt.YM  
US-09-475-515-80

Query Match 75.3%; Score 1858; DB 4; Length 2305;  
Best Local Similarity 89.0%; Pred. No. 5.3e-309;  
Matches 2058; Conservative 0; Mismatches 235; Indels 19; Gaps 4;

QY 170 GCGCAAGGAGGCGCACAGATGAAGACTGSCACGAGGCGCAGGCCAACTTCTTCGCG 229  
DB 1 GCGGCCCGAAGGACACCAATGAAGATTGCATGAGACAGCTAAATTTT-AGGG 59

QY 230 AGGACCTTGGCTTCCCCCAGGCGAAGGCCGCGAGTTCCCCAGGAGCAGAACCGGCCA 289  
DB 60 AAGATCTGGCTTCTTACAGGGAAGGCCAGGGAATTTCTTCAGAGCAGACGAGGCCA 119

QY 290 ACAGCCCCACAGCCGCGAGCTGAGGTGCGCGCGC-----ACAACCCCGCGAGCGAGG 343  
DB 120 ACAGCCCCACAGAGAGAGCTTCAAGTGTGGGAGGAGAAACAACTCCCTCTTCAGAAG 179

QY 344 CCGCGCGCGAGCGCCAGGCA-----CCCTGAACTTCCCCAGATCACCCTGTGGCAGC 397  
DB 180 CAGAGCCGATAGACAAGGAATGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239

QY 398 GCCCCTTGGTGAAGATCAAGGTGGCGGCCAGATCAAGGAGGCGCTCTGGACACCGCGC 457  
DB 240 GACCCCTCGTACAAATGAAGATCGGGGGCACTCAAGGAAGCGCTCTGATACAGGAG 299

QY 458 CGCAGCACCGTCTCGAGAGATGAGCTTCCCGCGCAAGTGAAGCCCAAGATGATCG 517  
DB 300 CAGATGATACAGTATTAGAAGAAATGAATTTGCGAGAAATGGAACCAAAATGATAG 359

QY 518 GCGCATCGCGGCTTCAATCAAGTGGCGCAGTACAGCAGATCCTGATCGAGATCTGCG 577  
DB 360 GGGGATCGGGGCTTCAATCAAGTGAAGCAGTACGACCACTCTGTAGAAATCTGTG 419

QY 578 GCAAGAGGCGCATCGGCACCGTGTGATCGGCGCCCAACCGCGTGAACATCATCGCGCGCA 637  
DB 420 GACATAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAA 479

QY 638 ACATGCTGACCCAGCTGGGTGCACTTCACTGAACTTCCCATCAGCCCCCATCGAGACCTGC 697  
DB 480 ATCTGTTGACCCAGATCGGCTGCACTTGAATTTCCCATCAGCCCCCATTTGAGACGCTGC 539

QY 698 CCGTGAAGCTGAGCCCGCATGAGCGCCCGCAGGTGAAGCAGTGGCCCTGACCGAGG 757  
DB 540 CCGTGAAGTGAAGCCGGGATGAGCGCCCGCAGGTGAAGCAGTGGCCATTTGACCGAGG 599

QY 758 AGAAGATCAAGGCGCTTACCGCATCTGCGAGGAGATGGAGAGGAGGCGCAAGATCACCA 817  
DB 600 AGAAGATCAAGGCGCTTGTGAGATCTGACCGGAGATGGAGAGGAGGCGCAAGATCAGCA 659

QY 818 AGATCGCGCCCGAGAACCCCTTACAAACACCCCGTGTTCGCGCATCAAGAAAGAGCAGCA 877  
DB 660 AGATCGCGCCCGAGAACCCCTTACAAACACCCCGTGTTCGCGCATCAAGAAAGAGCAGCA 719

QY 878 CCAAGTGGCGCAAGCTTGTGAGCTTCCGCGAGTGAACAAAGCGCAGCCAGGACTTCTGGG 937  
DB 720 CCAAGTGGCGCAAGCTTGTGAGCTTCCGCGAGTGAACAAAGCGCAGCCAGGACTTCTGGG 779

QY 938 AGGTGACGTGGGCATCCCCACCCCGCGCTTGAAGAAAGAAAGAGAGCTGACCGTGC 997  
DB 780 AGGTGACGTGGGCATCCCCACCCCGCGCTTGAAGAAAGAAAGAGAGCTGACCGTGC 839

QY 998 TGGAGCTGGGCGAGCGCTTCTTCAAGTGGCGCTTGAAGAGGAGCTTCCGAAAGTACACCG 1057  
DB 840 TGGAGCTGGGCGAGCGCTTCTTCAAGTGGCGCTTGAAGAGGAGCTTCCGAAAGTACACCG 899

QY 1058 CTTTACCATCCCCCAGCATCAACAAAGAGAGAGCGCGCATCCGCTACAGTACAAAGTGC 1117  
DB 900 CTTTACCATCCCCCAGCATCAACAAAGAGAGAGCGCGCATCCGCTACAGTACAAAGTGC 959

QY 1118 TGCCCCAGGCTGGAAGGCGAGGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGG 1177  
DB 960 TGCCCCAGGCTGGAAGGCGAGGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTTGG 1019

QY 1178 AGCCCTTCCGCGCGCGCAAGCGAGATCTGATCTTACCATGATGAGAGCAGCTGTACG 1237  
DB 1020 AGCCCTTCCGCAAGAGCAAGCGAGATCTGATCTTACCA-----GGCCCCCTGTACG 1073

QY 1238 TGGGAGCGACCTGGAGATCGGCGCAGCAGCGCGCAAGATCGAGAGCTGCGCAAGCACC 1297  
DB 1074 TGGGAGCGACCTGGAGATCGGCGCAGCAGCGCGCAAGATCGAGAGCTGCGCAAGCACC 1133

QY 1298 TGTGCGCTGGGCTTCAACACCCCGCAAGAGCAAGCAAGAGAGCGCCCTTCTCTGT 1357  
DB 1134 TGTGCGCTGGGCTTCAACACCCCGCAAGAGCAAGCAAGAGAGCGCCCTTCTCTGT 1193

QY 1358 GATGGCTAGGAGCTGACCCCGCAGAGTGGAGCGCTGAGCCCATCGAGTGTGCCGAGA 1417  
DB 1194 GATGGCTAGGAGCTGACCCCGCAGAGTGGAGCGCTGAGCCCATCATGCTGCCCGAGA 1253

QY 1418 AGCAGAGCTGCAACCGTGAACGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGCAGCC 1477  
DB 1254 AGCAGCTGCAACCGTGAACGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGCAGCC 1313

QY 1478 AGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGGCGCCCAAGGCCC 1537  
DB 1314 AGATCTACCGCGCATCAAGTGAAGCAGCTGTGCAAGCTGTGCGGCGCAGGAGGCC 1373

QY 1538 TGACCGACATCTGCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAGACCGCGAGA 1597  
DB 1374 TGACCGAGGTGATCCCCCTGACCGAGGAGCGCGAGCTGGGCGCAGAACCGCGAGA 1433

QY 1598 TCCTGCGCAGCGCTGACCGCGTGTACTACGACCCAGCAGGAGCTGTGTGGCGCCGAGA 1657  
DB 1434 TCCTGAAGAGCCCGTGCACGAGGTGTACTACGACCCAGCAGGAGAGCTGTGTGGCGCGAGA 1493

QY 1658 TCAGAGAGCGGCGCACGACAGTGGACCTTACAGATCTTACAGGAGCCCTTCAAGAAC 1717  
DB 1494 TCAGAGAGCGGCGCACGAGGCGAGTGGACCTTACAGATCTTACAGGAGCCCTTCAAGAAC 1553

QY 1718 TGAAGAGCGCAAGTACGCGCAAGTGGCGACCGCGCCACCAACAGAGCTGAGCAGCTGA 1777  
DB 1554 TGAAGAGCGCAAGTACGCGCGATCGCGCGCGCGCCACCAACAGAGCTGAGCAGCTGA 1613

QY 1778 CCGAGCGCTGCAAGAGATCGCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT 1837  
DB 1614 CCGAGCGCTGCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1673

QY 1838 TCCGCTTCCCATCCAGAGAGGAGACCTTGGAGAGCCTTGGTGAAGCCGACTTCTGGCAGGCCA 1897  
DB 1674 TCAAGCTTCCCATCCAGAGAGGAGACCTTGGAGAGCCTTGGTGAAGTGAAGTGAAGTGAAGT 1733

QY 1898 CTTGATCCCGAGTGGAGTGTGTGAACACCCCGCGCTTGGTGAAGCTGTGTGAGCAGC 1957  
DB 1734 CTTGATCCCGAGTGGAGTGTGTGAACACCCCGCGCTTGGTGAAGCTGTGTGAGCAGC 1793

QY 1958 TGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCCCAAGCGCG 2017  
DB 1794 TGGAGAGGAGCCCATCGTGGCGCGCGAGACCTTCTACGTGGAGCGCGCGCCCAAGCGCG 1853



Qy 1358 GGATGGGCTACGAGTGCACCCCGACCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGA 1417  
Db |||||  
1194 CCAT-----CGAGCTGCACCCCGACCAAGTGGACCGTGCAGCCCATCGTGCAGCCGAGA 1247  
Qy 1418 AGGAGAGCTGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAATCGGGCCAGCC 1477  
Db |||||  
1248 AGGACAGCTGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAATCGGGCCAGCC 1307  
Qy 1478 AGATCTACCCCGGCATCAAGTGGCCGACGCTGTGCAAGCTCTCTGGGGCCCAAGGCC 1537  
Db |||||  
1308 AGATCTACCCCGGCATCAAGTGAAGCAGCTGTGCAAGCTCTCTGGGGCCCAAGGCC 1367  
Qy 1538 TGACCGACATCGTGCCTTGAACGAGAGGCGGAGCTGGAGCTGGCGGCGAGAACCGCGAGA 1597  
Db |||||  
1368 TGACCGAGGTGATCCCTGACCGAGAGGCGGAGCTGGAGCTGGCGGCGAGAACCGCGAGA 1427  
Qy 1598 TCCTGCGGAGCCGCTGACCGGCTGTACTACGACCCCGAGCAAGACCTGTGTGCGCAGA 1657  
Db |||||  
1428 TCCTGAAGGAGCCGCTGACGAGGTGACTACGACCCCGAGCAAGGACCTGTGTGCGCAGA 1487  
Qy 1658 TCCAGAGAGCGGCGCACGACGAGTGGACCTTACAGATCTACGAGAGCCCTTCAAGAAC 1717  
Db |||||  
1488 TCCAGAGAGCGGCGCACGAGGCGAGTGGACCTTACAGATCTACGAGAGCCCTTCAAGAAC 1547  
Qy 1718 TGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCCCAACCAACGACGTGAAGCAGCTGA 1777  
Db |||||  
1548 TGAAGACCGGCAAGTACGCGCGCATGCGCGCGCCCAACCAACGAGCTGAAGCAGCTGA 1607  
Qy 1778 CCGAGGCGGTGCAGAAATCGCCATGAGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGT 1837  
Db |||||  
1608 CCGAGGCGGTGCAGAAATGAGCACCGGAGGACATCGTGTATCTGGGGCAAGATCCCAAGT 1667  
Qy 1838 TCCGCTCCGCGCATCCAGAGGAGACCTGGGAGACCTGTGTGACCGACTACTTGGCGAGGCCA 1897  
Db |||||  
1668 TCAAGCTGCCCATCCAGAGAGAGACCTGGGAGGCTGTGTGATGAGTACTTGGCGAGGCCA 1727  
Qy 1898 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTGACAGC 1957  
Db |||||  
1728 CTTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTGACAGC 1787  
Qy 1958 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTAGCTGGAGCGGCGCGCCCAACCGCG 2017  
Db |||||  
1788 TGGAGAGGAGCCCATCATCGTGGCGCGAGACCTTCTAGCTGGAGCGGCGCGCCCAACCGCG 1847  
Qy 2018 AGACCAAGTCCGCAAGCGCGCTACGTGACCGGAGCGGCGCGGCGGCGGAGATCGTGAAGC 2077  
Db |||||  
1848 AGACCAAGTCCGCAAGCGCGCTACGTGACCGGAGCGGCGCGGCGGCGGAGATCGTGAAGC 1907  
Qy 2078 TGACCGGAGACCCACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCGAGGACA 2137  
Db |||||  
1908 TCGCGGACACACCAACCAAGAGACCGAGCTGCAGGCGCATCCACCTGGCCCTGCGAGGACA 1967  
Qy 2138 CGGCGAGCGAGGTGAACATCGTGAACGACGCGAGTGCAGGCGCATCCAGGCGCC 2197  
Db |||||  
1968 CGGCGCTGGAGGTGAACATCGTGAACGACGCGAGTGCAGGCGCATCCAGGCGCC 2027  
Qy 2198 AGCCCGACAGAGCGAGCGAGCTGTGAAACGAGATCATCGAGCAGCTGATCAAGAAGG 2257  
Db |||||  
2028 AGCCCGACAGAGCGAGCGAGCTGTGAAACGAGATCATCGAGCAGCTGATCAAGAAGG 2087  
Qy 2258 AGAAGGTGTACCTGAGCTGGGTGCGCCCGACCAAGGCGCATCGGCGGCAAGCAGATCG 2317  
Db |||||  
2088 AGAAGGTGTACCTGGCTGGGTGCGCCCGACCAAGGCGCATCGGCGGCAAGCAGGTGG 2147  
Qy 2318 ACAAGCTGGTGAACAGGCGATCCGCAAGGTGTGTTCTTGGAGCGGATCGATGGCGGCA 2377  
Db |||||  
2148 ACAAGCTGGTGAACAGGCGATCCGCAAGGTGTGTTCTTGAACGCGATCGATGGCGGCA 2207  
Qy 2378 TCGTGTACTACAGTACATGACGACCTGTACGTGGGCGAGCGGCGGCGCTAGGATCGATT 2437  
Db |||||  
2208 TCGTGTACTACAGTACATGACGACCTGTACGTGGGCGAGCGGCGGCGCTAGGATCGATT 2267

Qy 2438 AAAAGCTTCCGGGGCTAGCACCCGCTGAATTC 2469  
Db |||||  
2268 AAAAGCTTCCGGGGCTAGCACCCGCTGAATTC 2299

## RESULT 7

US-09-552-950-2  
; Sequence 2, Application US/09552950  
; Patent No. 6541248  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; TITLE OF INVENTION: Anti-Viral Vectors  
; FILE REFERENCE: 674524-2004  
; CURRENT APPLICATION NUMBER: US/09/552,950  
; CURRENT FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4307  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence.gagpol - codon  
; OTHER INFORMATION: optimised gagpol sequence  
US-09-552-950-2

Query Match 68.0%; Score 1678.6; DB 4; Length 4307;  
Best Local Similarity 82.9%; Pred. No. 2.8e-278;  
Matches 1968; Conservative 0; Mismatches 389; Indels 16; Gaps 4;

Qy 12 CATGCCGAGGCCATGAGCAGG---CCACGAGGCCAAATCTCTGATGACGAGCGCA 68  
Db 1086 CTTGGCTGAGGCCATGAGCAGGTCACCACTCCGCTACCATCATGATGACGAGCGGCA 1145  
Qy 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGGAGGCCACATCGC 128  
Db 1146 CTTTCGGAACAACAGCAAGATCGTCAAGTCTTCAACTGTGTCGCAAGAGGGCACACGC 1205  
Qy 129 CCGCAACTGCGGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCAGGCAAGGAGGCCACCA 188  
Db 1206 CCGCAACTGCGGCGCCCTAGGAAAGAGGGCTGTGGAATGCGGCAAGGAGGCCACCA 1265  
Qy 189 GATGAAGACTGCACCGAGCGCCAGGCAACTCTTCCGCGAGGACCTTGGCTTCCGCCA 248  
Db 1266 GATGAAGACTGTACGAGAGACAGGCTAA-TTTTGTAGGAAAGATCTGGCTTCTCTACA 1324  
Qy 249 GGGCAAGGCGCGAGTTCCTCCAGCGAGAGAACCGCGCCCAACAGCCCAACCGCCGGA 308  
Db 1325 AGGGAAGGCCAGGGAATTTTCTTCAAGAGCAGACCCGCGCAACAGCCCAACCGCCGGA 1384  
Qy 309 GCTGCAGGT-----GCGCGCGACAAACCCCGCAGCGAGGCGCGCGCCAGCGCCAGGG 362  
Db 1385 GCTTCAGTCTGGGGTCCGACAACTCTCCCTTCGAGAGCAGGAGCGGCGCCAGGG 1444  
Qy 363 CA-----CCCTGAATTTCCCGAGATCACCTGTGTGACGCGCCCTGTGTGAGCATCAA 416  
Db 1445 CACGCTGCTTCAACTTCTCTCAGGTACGCTTGGCAGCGACCCCTCTGTCAACATCAA 1504  
Qy 417 GGTGGGCGGCGAGATCAAGGAGGCTGTGGAACACCGGCGCGCCAGCACACCGTGTGGA 476  
Db 1505 GATCGGGGGGCGAGTCAAGGAGGCTCTCTTGGACACCGGAGCAGACACCGTGTGGA 1564  
Qy 477 GGAGATGAGCTGCTCCCGCAAGTGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536  
Db 1565 GGAGATGCTGTTGCCAGCGCTGGAAGCCGAGATGATCGGGGAATCGGCGGTTTCAT 1624  
Qy 537 CAAGGTGCGCAGTACGACAGATCTCTGATCGAGATCTCGGCGCAAGAGGCATCGGCAC 596  
Db 1625 CAAGGTGCGCAGTATGACCAAGATCTCTCATCGAAATCTGCGGGCCCAACAGGCTATCGGTAC 1684  
Qy 597 CGTGTCTATCGGCGCCACCCCGTGAACATCATCGGCGGCAACATGTGTGACCCAGCTGGG 656  
Db 1685 CGTGTGTGGGGCCCCACACCCCGTCAACATCATCGGAGCGCAACCTGTGTGACGAGATCGG 1744

Qy	657	CTGCACCTTGAACCTTCCCATTCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGG	716
Db	1745	TTGCACGCTGAACCTTCCCATTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG	1804
Qy	717	CATGGACGGCCCAAGGTGAAGCTGGCCCTCTACCGCAGGAGNAGATCAAGGCCCTGAC	776
Db	1805	GATGGACGGCCCGAAGGTCAAGCAATGGCCATTGTACAGAGAGGAGAATCAAGGCATCTGGT	1864
Qy	777	CGCCATCTGCGAGGAGATGGAAGAAGGAGGCAAGATCACCAAGATCGGCCCGCAGAACCC	836
Db	1865	GGAGATTGACAGAGATGGAAAGGAAGGNAATCTCCAGATTGGGCTTGNAAACC	1924
Qy	837	CTACAAACACCCCGGTGTTCCGCATCAAGAAAGGAAGGACAGCACCAAGTGGCGCAAGCTGGT	896
Db	1925	GTAACAAACACCCCGGTGTTCCCAATCAAGAAAGGAAGGACTCGACGCAAAATGGCGCAAGCTGGT	1984
Qy	897	GGACTTCCGGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCAATCCC	956
Db	1985	GGACTTCCGGAGCTGAACAGCGCACCGCAAGGACTTCTGGGAGGTTCAGCTGGGCAATCCC	2044
Qy	957	CCACCCCGCGGCTGAAGAAGAAAGAGCGTGACCGTGTGGAAGTGGCGGCGACGCGTA	1016
Db	2045	GCACCCCGCAGGGCTGAAGAAGAAAGAAATCCGTGACCGTACTGGATGGGTGATGCGTA	2104
Qy	1017	CTTCAGCGTGCCCTTGACGAGGAGACTTCCGGAAGTACACCGCTTCAACATCCCCAGCAT	1076
Db	2105	CTTCTCCGTTCCCTTGACGAAAGACTTCAGGAAGTACACTGCGCTTCAAAATCCCTTCGAT	2164
Qy	1077	CACAACGAGACCCCGGCAATCCGCTACCACTACAAGCTGCTGCCCAGAGGCTGGAAAGG	1136
Db	2165	CAACAACGAGACACCGGGATTCGATATCAGTACAAAGCTGCTGCCCAGGGCTGGAAAGG	2224
Qy	1137	CAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAA	1196
Db	2225	CTCTCCCGCAATCTTCCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCGCAAAACAGAA	2284
Qy	1197	CCCCGAGATCGTGATCTACACGATACATGAGAGACTGTATCGTGGGCAAGCACTTGAGAT	1256
Db	2285	CCCCGACATCGTCTATCTATCAGTACATGAGATGACTTGTAGCTGGGCTCTGATCTAGAGAT	2344
Qy	1257	CGGCCAGCACCGCGCCAAGATCGAGGAGCTGCGCAAGCACTGTGCGCTGGGGCTTCAC	1316
Db	2345	AGGGCAGCACCGCAACCAAGATCGHAGAGCTGCGCCAGACCTGTTGAGTTGGGACGTGAC	2404
Qy	1317	CACCCCGCAAGAAGCACCAAGAGAGCCCCCTCTCTGTGGATGGGCTACAGAGTGA	1376
Db	2405	CACACCGCAAGAAGCACCAAGAGAGCTCTCCCTTCTCTGGATGGGTTACGAGCTGA	2464
Qy	1377	CCCGGACAAAGTGAGCCGTGCAGCCCATCGAGCTGCCCCGAGAGGAGAGCTGGAACCTGAA	1436
Db	2465	CCCTGCAAAATGGACCGTGCAGCCATCTGTGCTGCCAGAGAAAGACAGCTTGGATCTGCAA	2524
Qy	1437	CGACATCCAGAAGCTGTGTGGGCAAGCTGAATGGGGCCAGCCAGATCTACCCCGGCATCAA	1496
Db	2525	CGACATACAGAGCTGTGTGGGAAGTTGAATGGGCCAGTCAAGTTTACCCAGGATTA	2584
Qy	1497	GGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCAGGCCCTGACCGACATCTGTGCCCT	1556
Db	2585	GGTGAGCGACGTGTGCAAACTCTCTCCCGGAAACCAAGGCCACTCACAGGTTGATCCCCCT	2644
Qy	1557	GACCGAGAGCCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCCTGCGCGAGCCCGTGCA	1616
Db	2645	AACCGAGAGCCGAGCTCGAACTGGCAGAAAAACCGAGAGATCTTAAGAGAGCCCGTGCA	2704
Qy	1617	CGGCGTGTACTACAGCCCCCAGCAGGACCTGGTGGCCCGAGATCCAGAAGCAGAGGCCACGA	1676
Db	2705	CGGCGTGTACTAGACCCCTTCAAGGACCTGATCGCCGAGATCCAGAAGCAGGGGCAAGG	2764
Qy	1677	CCAGTGACCTTACAGATCTTACAGGAGCCCTTCAAGAACTTGAAGACCCGCAAGTACGC	1736
Db	2765	CCAGTGACCTTACAGATTTTACAGGAGCCCTTCAAGAACTTGAAGACCCGCAAGTACGC	2824

Qy	1737	CAAGATCGCAGCGCCACACCAACGACGTGAAGACAGCTACCGAGCCCGTCGAGAAGAT	1799
Db	2825	CCGGATGAGGGGTGCCACACTAACACGCTCAAGCAGCTGACCGAGGCGGTGCGAGAAGAT	2884
Qy	1797	CGCCATCGAGAGCATCGTGATCTGGGCGAAGACCCCAAGCTTCGCGCTGCCCATCCAGAA	1856
Db	2885	CACACCCGAAGCATCGTGATCTGGGGAAGAATCTCTAAGTTCAAGCTGCCCATCCAGAA	2944
Qy	1857	GGAGACTTGGGAGACCTTGGTGGACCGACCTAATCGGCAGGCCACTTGGATCCCCAGTGGGA	1916
Db	2945	GGAAACTCTGGGAAACTTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCTCGAGTGGGA	3004
Qy	1917	GTTTCGTGAACACCCCCCCTCGTGAAGCTGGTACCAAGCTGGAGGAAGAGGCCATCAT	1976
Db	3005	GTTTCGTCAACACCCCTCCCTCGTGAAGCTGGTACCAAGCTGGAGGAAGAGGCCATAGT	3064
Qy	1977	CGGCGCCGAGACCTTCTACCTGGACGGCGCGGCCAACCGCGAGACCAAGATCGCGCAAGGC	2036
Db	3065	GGGCGCCGAAACCTTCTACGTGGATGGGCGCGCTAACAGGAGACTTAAGCTGGGCAAAAGC	3124
Qy	2037	CGGCTACGTGACCGGACCGGGGCGGCGAAGATGCTGAGGCTGACCGAGACCAACCAACCA	2096
Db	3125	CGGATACGTCACTAAACGGGGCGAGACAGAAAGTTGTCAACCCTCACTGACACCAACCAACCA	3184
Qy	2097	GAGACCGAGCTGACAGCCCATCAGCTGGCCCTGCGAGGACAGCGGCAGCGAGGTGAACAT	2156
Db	3185	GAGACTGAGCTGCAGGCGCAATTAACCTCGCTTTGAGGATCTCGGCTCGAGGTGAACAT	3244
Qy	2157	CGTGACCGACAGCGAGTACGACCTTGGGCACTATCCAGGCCCCAGCCGACAAAGAGCGAGAG	2216
Db	3245	CGTGACAGACTCTCAGTATGCCCTTGGGCACTATTCAAGCCCGAGCCAGAGGTGATGTC	3304
Qy	2217	CGAGCTGGTGAACGAGATCATTCGACGACGTGATCAGAAGGAGAAAGTGTACCTGAGCTG	2276
Db	3305	CGAGCTGGTCAATCAGATCATTCGACGACGTGATCAAGAGGAAAGGTCTATCTGGGCTG	3364
Qy	2277	GGTGCCCGCCCAAGGGCATCGGCGCAACAGCAGCAGATCGCAAGCTGGTGAGCAAGGG	2336
Db	3365	GGTACCCGCCCAAGAGGCAATTGGCGCATGAGCAGGTGCAACAGCTGCTCGGCTGG	3424
Qy	2337	CATCCGCAAGGTGCTGTTCTCTGGAACGCACTCGA	2369
Db	3425	CATCAGAAAGGTGCTATTCTCTGATGGCATCGA	3457

```

RESULT 8
US-09-936-572-2
; Sequence 2, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936,572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: gagpol-Syngp-codon optimised gagpol sequence
US-09-936-572-2

```

Query Match	67.2%;	Score 1659.4;	DB 4;	Length 4307;
Best Local Similarity	82.4%;	Pred. No. 5.3e-275;		

Query Match	67.2%;	Score 1659.4;	DB 4;	Length 4307;
Best Local Similarity	82.4%;	Pred. No. 5.3e-275;		



Matches 1956; Conservative 0; Mismatches 401; Indels 16; Gaps 4;			
Qy	12	CATGCCGAGCCATGAGCCAGG---CCACAGGCGCAACATCTGTATGACGCGCAGCAA	68
Db	1086	CCTGGCTGAGGCCATGAGCCAGGTGACCACTCCGCTACCATCATGATGACGCGCGCAA	1145
Qy	69	CTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCCACATCGC	128
Db	1146	CTTTCCGAAACAACGCAAGATCGTCAAGTGTCTTCAACTGTGTCGAAGAAGGGCACAGC	1205
Qy	129	CCGCAACTGCCCGCCCCCGCAAGAAGGGGTGCTGGAAGTGCAGGCGCAAGGAGGCCACCA	188
Db	1206	CCGCAACTGACAGGCCCTTAGGAAGAAGGGGTGCTGGAATGCGGCAAGGAGGCCACCA	1255
Qy	189	GATGAAGGATGACACAGAGGCCAGGCCAACTTCTCCGAGAGACTGTGCGCTTCCGCCA	248
Db	1266	GATGAAGGATGTACTAGAGACAGGGCTAA-TTTTTTAGGGAAGATCTGGCGCTTCTTACA	1324
Qy	249	GGGCAAGGCCCGAGTTCCCGAGGAGCAGAAACCGCGCCAAACAGCCCCACCGCCGGA	308
Db	1325	AGGGAAGGCCAGGGAAATTTTCTTCAGAGCAGACAGAGCCCAACAGCCCTCCAGAGAGA	1384
Qy	309	GCTGACGGTGGCGG-----CGACAACCCCGCAGCAGGCGCGCGCCGAGCCCAAGG	362
Db	1385	GCTTCAGGTCTGGGTAGAGNACAACACTCCCCCTCAGAGCAGAGCCGATAGACAAG	1444
Qy	363	CA-----CCCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGATCAAA	416
Db	1445	AACGTGATCTTTAACTTCCCTCAGATCACTCTTTTGGCAAGCACCCCTCGTCACAATAA	1504
Qy	417	GGTGGGCGCCAGATCAAGAGGCGCTGTGGACACCGGCGCCGACAGACACCGTGTGGA	476
Db	1505	GATAGGGGGGAGCTCAAGGAGGCTCTCTTGGAACACCGGAGCAGACACCGTGTGGA	1564
Qy	477	GGAGATGAGCTGCCCGCAAGTGGAGCCCAAGATGATCGGCGCATCGGCGCTTCAT	536
Db	1565	GGAGATGCTGTGCGACGCGCTGGAAGCCCAAGATGATCGGCGGAATCGGCGGTTTCAT	1624
Qy	537	CAAGGTGCGCAGTACGACAGATCTCTGATCGAGATCTGGGCAAGAGGCCATCGGCAC	596
Db	1625	CAAGGTGCGCCAGTATGACAGATCTCTCATCGAATCTGCGGCCACAGGCTATCGGTAC	1684
Qy	597	CGTGTGATCGGCCCGCCACCCCGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGG	656
Db	1685	CGTGTGTTGGCGGCCACACCCCGTCAACATCATCGAGCGCAACCTGTTGACGCAAGTCGG	1744
Qy	657	CTGACCCCTGAACTTCCCGCATCAGCCCATCGACCGTGCCTGGAAGCTGGAAGCCCGG	716
Db	1745	TTGACCGCTGAATTTCCCGATTTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG	1804
Qy	717	CATGGACGGCCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGAGAGAAATCAAGGCCCTGAC	776
Db	1805	GATGGACGGCCGAGGTCAAGCAATGSCCATTCACAGAGGAGAAATCAAGGCACTGGT	1864
Qy	777	CGCATCTCGAGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCCCCGAGAACCC	836
Db	1865	GGAGATTTGCACAGATGGAAGAAGGAGGAAATCTCCAAGATTGGGCGCTGAGAACCC	1924
Qy	837	CTACAACACCCCGTGTTCGCAATCAAGAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGT	896
Db	1925	GTACAACACCCCGTGTTCGCAATCAAGAAGAGGACTCGACGAAATGGGCGAAGCTGGT	1984
Qy	897	GGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGCAATCCC	956
Db	1985	GGACTTCCGCGAGCTGAACAGCGCACCGAAGACTTCTGGGAGGTTCAGCTGGGCAATCCC	2044
Qy	957	CCACCCCGCGGCTGAAGAAGAGAGAGCGTGACCCGTGTGACCGTGGCGCGACGCTTA	1016
Db	2045	GCACCCCGCGAGGCTGAAGAAGAGAGAAATCCGTGACCGTACTGAGTGGGTGATGCTTA	2104
Qy	1017	CTTCAGGCGTCCCTGACGAGGACTTCCGCAAGTACACCGCCTTTCACCATCCCGACAT	1076
Db	2105	CTTCTCCGTTTCCCTGGAGGAAGACTTTCAGGAAGTACACTGCGCTTTCACATCCCTTCGAT	2164

Qy	1077	CAACAACGAGACCCCCCGGCATCCGCTACCAAGTACAACGTCGTGCTCCCAAGGCTGGAAGGG	1136
Db	2165	CAACAACGAGACACCGGGGATTCGATATCAGTACAACGTCGTGCTCCCAAGGCTGGAAGGG	2224
Qy	1137	CAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGGGCGCCGCAA	1196
Db	2225	CTCTCCGCAATCTTCAGAGTAGCATGACCAAAATCCTGGAGCTTTCCGCAACAGAA	2284
Qy	1197	CCCCGAGATCGTGATCTACCAAGTACATGGAGCCTGTACGTGGGCGAGCAGCTGGAGAT	1256
Db	2285	CCCCGACATCGTCATCTATCAGTACATGGATGACTTGTACGTGGGCTCTGATCTAGAGAT	2344
Qy	1257	CGGCACGACCGCGCCCAAGATCGAGAGAGCTGCGCAAGGACCTGCTGGCTGGGGCTTCAC	1316
Db	2345	AGGCGACACCGCACCAAGATCGAGGAGCTCGCCAGCACCTGTTGAGGTGGGACTGAC	2404
Qy	1317	CACCCCGCACAAGAACCAACAGAGGAGCCCGCTCTCTGATGGATGGGCTACGAGCTGCA	1376
Db	2405	CACACCGCACAAGAACCAACAGAGGAGCCCTCTCTCTGGATGGGTTACGAGCTGCA	2464
Qy	1377	CCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGGAGAGCTGGACCGTGA	1436
Db	2465	CCCTGACAAATGGACCGTGCAGCCTATCGTCTGCCAGAGAAAGACAGCTGGACTGTCAA	2524
Qy	1437	CGACATCCAGAAGCTGTGTGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGATCAA	1496
Db	2525	CGACATACAGAAGCTGTGTGGCAAGTTGAACTGGGCGCAGTCAGATTTACCCAGGATTA	2584
Qy	1497	GGTGGCGCAGCTGTGCAAGCTGTGCGGGGCGCAAGCCCTGACCGACATCGTGGCCCT	1556
Db	2585	GGTGGAGCAGCTGTGCAAACTCTCTCCGCGAAACCAAGGCACTCACAGAGGTGATCCCGCT	2644
Qy	1557	GACCGAGAGCGCGAGCTGAGCTGGCGGAGAACCGCGAGATCTCTGCGAGGCGCGTGA	1616
Db	2645	AACGAGAGCGCGAGCTCGAACTGGCGAGAAACCGGAGAGATCTCTAAGGAGCCCGTGA	2704
Qy	1617	CGGCGTGTACTACGACCCCGCAGCAAGGACCTGTGTGGCCGAGATCCAGAGCAGGCGCA	1676
Db	2705	CGGCGTGTACTATGACCCCTCCAGGAGCTGTGCGCGAGATCCAGAGCAGGCGGCAAGG	2764
Qy	1677	CGAGTGGACCTACAGATCTACCGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTAGCG	1736
Db	2765	CGAGTGGACCTATCAGATTTTACAGAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTAGCG	2824
Qy	1737	CAAGATCGCACCGCCGACACCAACGACGCTGAAGCAGCTGACCGAGCGCGTGCAGAAAGAT	1796
Db	2825	CCGGATGAGGGGTGCCCACTAACGACGCTCAGCAGCTGACCGAGGCGCGTGCAGAAAGAT	2884
Qy	1797	CGCCATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCGCCATCCAGAA	1856
Db	2885	CACCACCGAAAGCATCTGTATCTGGGGAAGAGCTCTTAAGTTCAAGCTGCGCCATCCAGAA	2944
Qy	1857	GGAGACTGGAGACCTGTGTGGACCGACTACTGGCAGGCGCACCTGGATCCCGGAGTGGGA	1916
Db	2945	GGAAACCTGGGAACTCTGTGTGGACAGAGTATGGCAGGCGCACCTGGATTCCTGAGTGGGA	3004
Qy	1917	GTTTCGTGAACACACCCCGCTGTGTGAAGCTGTGTGTACAGCTGAGAGAGGAGGCCATCAT	1976
Db	3005	GTTTCGTCAACACCCCTCTCCCTGTGTGAAGCTGTGTGTACAGCTGAGAGAGGAGGCCATAGT	3064
Qy	1977	CGGCGCGCAGACCTTCTACGTGGAACCGCGCGCCCAACCGCGAGACCAAGATTCGCAAGGC	2036
Db	3065	GGGCGCGCAACCTTCTACGTGGAACCGCGCGCCCTAAACAGGAGACTAAGCTGGGCAAGGC	3124
Qy	2037	CGGCTAGCTGACGACCGGGGCGGCAAGATGCTGAGGCTGTGACCGAGACCAACCAACCA	2096
Db	3125	CGGATACGTCACTAAACCGGGGCGACAGAAAGTGTTCACCTCTCACTGACCAACCAACCA	3184
Qy	2097	GAAGACCGAGCTCGAGGCCATCCAGCTGGCCCTTGCAGGACAGCGGAGCGGAGTGAACAT	2156
Db	3185	GAAGACTGAGCTGAGGCCATTTACCTCGCTTTTGCAGGACTCGGCGCTTGGAGGTGAACAT	3244

QY 2157 CGTGACCGACGAGCCAGTACGCTGGGCGATCATCCAGGCGCCAGCCGCAAGACGAGAG 2216  
 DB 3245 CGTGACAGAGCTCTCAGTATGCCCTGGGCATCATTTCAAGCCAGCCAGACGAGTGGTC 3304  
 QY 2217 CGAGCTGGTGAACACGATCATCAGACGAGCTGTATCAAGAGGAGAAAGGTGTACCTGAGCTG 2276  
 DB 3305 CGAGCTGGTCAATCAGATCATCAGACGAGCTGTATCAAGAGGAGAAAGGTCTATCTGGCCTG 3364  
 QY 2277 GGTGCGCGCCCAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGG 2336  
 DB 3365 GGTACCGCGCCCAAGAGGCAATTGGCGCAATGAGCAGGTGCAACAAGCTGGTCTCGGCTGG 3424  
 QY 2337 CATCCGCAAGGTGCTGTCTTCGACGCGCATCGA 2369  
 DB 3425 CATCAGAGGTGCTATTCTCGATGGCATCGA 3457

RESULT 9

US-09-936-572-14  
 ; Sequence 14, Application US/09936572  
 ; Patent No. 6783981  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UDEN, MARK  
 ; APPLICANT: MITROPHANOUS, KYRIACOS  
 ; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
 ; FILE REFERENCE: 078983/0137  
 ; CURRENT APPLICATION NUMBER: US/09/936, 572  
 ; CURRENT FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: GB 990617.2  
 ; PRIOR FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 14  
 ; LENGTH: 4327  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: pSYNFP4-codon  
 ; OTHER INFORMATION: optimised HIV-1 gagpol with 20bp of the leader  
 ; OTHER INFORMATION: sequence of HIV-1  
 US-09-936-572-14

Query Match 66.9%; Score 1651.4; DB 4; Length 4327;  
 Best Local Similarity 82.2%; Pred. No. 1.2e-273;  
 Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;  
 QY 12 CATGGCGGAGGCCATGAGCCAGG---CCACCAGCGCCAAACATCTGTATGACGGCGAGCAA 68  
 DB 1106 CTGGCTGAGCCATGAGCCAGGTGACCACTCGGCTACCATCATGATGACGGCGGCAA 1165  
 QY 69 CTTCAAGGGCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGAGGGCCACATTCGC 128  
 DB 1166 CTTTCGNAACCAACGCAAGATCGTCAAGTGCTTCAACTGTGGCAAGAGGGCACACAGC 1225  
 QY 129 CCGCAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 198  
 DB 1226 CCGCAACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1285  
 QY 189 GATGAAGGACTGACACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 248  
 DB 1286 AATGAAGATGTGATGAGAGACAGGCTAA-TTTTTTAGGAAGATCTGGCCCTTCCACA 1344  
 QY 249 GGGCAAGGCGCGGAGTTCCTCCAGCGAGCAAGAACCGCGCCAAACAGGCCCAACGAGCGCGA 308  
 DB 1345 AGGGAAGGCGGAGGAAATTTCTTTCAGAGCAGACAGAGCCAAACAGGCCCAACGAGCAAGA 1404  
 QY 309 GCTGCGAGTGGCGG-----CGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 362  
 DB 1405 GCTTCAGGTTTGGGGAAGAGACAACTCCCTCTCAGAGGAGGAGCCCATAGACAAGG 1464  
 QY 363 CA-----CCCTGAACCTTCCCGCAGATCACCTGTGGCAGCGCGCCCTGTGAGCATCAA 416

DB 1465 AACTGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCAGCCCTCGTCAATAAA 1524  
 QY 417 GGTGGCGGCGCAGATCAAGAGGCGCTGCTGACACCGGCGCGAGACACCGTGTGGA 476  
 DB 1525 GATAGGGGGGAGCTCAAGGAGGCTCTCTTGACACCGGAGCAGACACCGTGTGGA 1584  
 QY 477 GAGATGAGCTGCGCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCGATCGGCGGCTTCAT 536  
 DB 1585 GAGATGTCTGTTGCGAGGCGCTGGGAAGCGAGATGATCGGCGGGAATCGGCGGTTTCAT 1644  
 QY 537 CAAGTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCGCATTCGGCAC 596  
 DB 1645 CAAGTGCGCCAGTATGACAGATCTCTCATCGAAATCTGCGGCCACAAGGCTATCGGTAC 1704  
 QY 597 CGTGTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 656  
 DB 1705 CGTGTGTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1764  
 QY 657 CTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 716  
 DB 1765 TTGCACGCTGAACTTCCCATTTAGCCCTATCGAGACGGTACCGTGAAGCTGAAGCGCGG 1824  
 QY 717 CATGACGCGCCCAAGGTGAAGCAGTGGCCCTTGAACGAGGAGAAAGATCAAGGCCCTTGAC 776  
 DB 1825 GATGACGCGCCCAAGGTCAAGCAATGGCCATTTGACAGAGGAGAAAGATCAAGGCACTGGT 1884  
 QY 777 CGCATCTCGGAGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 836  
 DB 1885 GGAGATTTTGCACAGAGATGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1944  
 QY 837 CTACAAACACCCCGCTGTTCGCCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 896  
 DB 1945 GTACAAACACCCCGCTGTTCGCCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2004  
 QY 897 GGACTTTCGCGAGCTGAAACAGGCGACCCAGAGACTTCTGGAGGAGTGCAGCTGGGCGATCCC 956  
 DB 2005 GGACTTTCGCGAGCTGAAACAGGCGACCCAGAGACTTCTGGAGGAGTGCAGCTGGGCGATCCC 2064  
 QY 957 CACCCCGCGGCGCTCAAG 1016  
 DB 2065 GCACCCCGGAGGCTGAAG 2124  
 QY 1017 CTTACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1076  
 DB 2125 CTTCTCGGTTCCCTGACGAGACTTTCAGAGAGTACACTGCTTCAAAATCCCTTCGAT 2184  
 QY 1077 CAACAAAGAGACCCCGGCGATCCGCTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 1136  
 DB 2185 CAACAAAGAGACCCCGGCGATCCGATATCAGTATCAGTATCAGTATCAGTATCAGTATCAGTATCAG 2244  
 QY 1137 CAGCCCGAGATCTTCCAGAGCAGCATGACCAAGATCCCTGGAGGCGCTTCCGCGCGCCGCA 1196  
 DB 2245 CTTCCCGCAATTTCCAGAGTACGATGACCAAAATCCCTGGAGGCGCTTCCGCGCGCCGCA 2304  
 QY 1197 CCGCGAGATCTGATCTTACCAAGTACGAGAGCTGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1256  
 DB 2305 CCGCGAGATCTGATCTTACCAAGTACGAGAGCTGTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2364  
 QY 1257 CCGCGAGCACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1316  
 DB 2365 AGGCGAGCACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2424  
 QY 1317 CACCCCGGAG 1376  
 DB 2425 CACACCGGAG 2484  
 QY 1377 CCGCGAG 1436  
 DB 2485 CCCTGACAAATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2544  
 QY 1437 CGACATCCAGAGAGCTGGTGGGCAAGCTGAACTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1496

Db 2545 CGACATACAGAAAGCTGGTGGGAAGTTGAACCTGGCCAGTCAGATTTACCCAGGATTA 2604  
Qy 1497 GGTGGCCAGCTGTCAAGCTGCTCGCGGGCCCAAGGCCCTGACCGACATCGTGGCCCT 1556  
Db 2605 GGTGAGGAGCTGTGCAAACTCTCCGCGGAACCAAGGCACTCAAGAGGTGATCCCTCT 2664  
Qy 1557 GACCGAGAGCCAGCTGGAGCTGGCCGGAACCGCGAGATCTTGGCGAGCCGCTGCA 1616  
Db 2665 AACCGAGAGCCAGCTGCAACTGGCAGAAACCGAGAGATCTTAAAGGAGCCGCTGCA 2724  
Qy 1617 CGGCTGTACTACACCCACCAAGGACCTGGTGGCCGAGATCCAGAAAGCAGGCGCCACGA 1676  
Db 2725 CGGCTGTACTATGACCCCTCAAGGACCTGATCGCCGAGATCCAGAAAGGCGGCAAG 2784  
Qy 1677 CCAGTGGACCTACAGATCTACAGAGGACCTTCAAGAACTGAAAGCCGCAAGTAGTC 1736  
Db 2785 CCAGTGGACCTACAGATTTACAGAGGACCTTCAAGAACTGAAAGCCGCAAGTAGTC 2844  
Qy 1737 CAAGATGGACCCGCCACCAACAGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAT 1796  
Db 2845 CCGATGAGGGGTGCCACACACTAAACAGCTCAAGCAGCTGACCGAGGCGGTGCAAGAT 2904  
Qy 1797 CGCATGAGAGCATCGTATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAA 1856  
Db 2905 CACCAACCGAAGCATCGTATCTGGGAAAGACTCTTAAGTTCAAGCTGCCCATCCAGAA 2964  
Qy 1857 GGAGACCTGGGAGACCTGGTGGACCGACTCTGGCAGGCCACCTGGATCCCGAGTGGGA 1916  
Db 2965 GGAAACCTGGGAACCTGGTGGACAGAGTATTGGCAGGCCACCTGGATTCTGAGTGGGA 3024  
Qy 1917 GTTGTGAAACCCCGCCCTGGTGAAGCTGTGTGATCCAGCTGGAGAGGAGCCCATCAT 1976  
Db 3025 GTTGTGAAACCCCGCCCTGGTGAAGCTGTGTGATCCAGCTGGAGAGGAGCCCATCAT 3084  
Qy 1977 CGGCGCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2036  
Db 3085 GGGCGCGGAACCTTCTACGTGGATGGGCGCGCTTAACAGGGAGACTAAGCTGGGCAAGC 3144  
Qy 2037 CGGTACGTGACCGACCGGGCGCGCAGAAAGATCGTGAGCCTGACCGAGACCAACCA 2096  
Db 3145 CGGATACGTCACTAACCGGGCGCAGACAGAGTGTGACCCCTCACTGACACCAACCA 3204  
Qy 2097 GAAGACGAGCTGAGGCGCATCAGCTGGCCCTGCGAGACAGCGGAGCGAGGTGAACAT 2156  
Db 3205 GAAGACTGAGCTGAGGCGCATTTACCTCGCTTTGCGAGCACTCGGCGCTGAGAGTGAACAT 3264  
Qy 2157 CGTGACGACGAGCAGTAAGCTGGGCGCATCATCCAGGCCCGCGCAGCAGAGCGAGAG 2216  
Db 3265 CGTGACAGACTCTCAGTATGCCCTGGGCGATCATTTCAAGGCCCGCGCAGCAGAGGTGATC 3324  
Qy 2217 CGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTG 2276  
Db 3325 CGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTATCTGCGCTG 3384  
Qy 2277 GGTGCGCCGCCAAGGGCATCGCGGCAACGAGCAGATGCAAGTGTGTGAGCAAGG 2336  
Db 3385 GGTACCGCGCCCAAGAGCATTGGCGGCAATGAGCAGGTGCAAGAGTGTCTCGGCTGG 3444  
Qy 2337 CATCGGCAAGGTGTGTCTGGAGCGGCATCGA 2369  
Db 3445 CATCAGGAAGGTGCTATTTCTGGATGTCATCGA 3477

## RESULT 10

US-09-936-572-13

; Sequence 13, Application US/09936572

; Patent No. 6783981

; GENERAL INFORMATION:

; APPLICANT: UDEN, MARK

; APPLICANT: MITROPHANOUS, KYRIACOS

; TITLE OF INVENTION: ANTI-VIRAL VECTORS

; FILE REFERENCE: 078863/0137

; CURRENT APPLICATION NUMBER: US/09/936,572

; CURRENT FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: GB 9906177.2  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4353  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PSYNGP3-codon  
; OTHER INFORMATION: optimised HIV-1 gapcol with leader sequence from  
; OTHER INFORMATION: the major splice donor  
US-09-936-572-13

Query Match 66.9%; Score 1651.4; DB 4; Length 4353;

Best Local Similarity 82.2%; Pred. No. 1.2e-273;

Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;

Qy 12 CATGCCGAGGCCATGAGCCAGG---CCACCAGCGCCAAACATCTGTATGTCAGCGCAGCAA 68  
Db 1132 CTTGGCTGAGGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGTGAGCGGGCAA 1191  
Qy 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCCACATCGC 128  
Db 1192 CTTTCGGAAACAACCGCAAGATCGTCAAGTGTCTCAACTGTGCAAGAAAGGCCACACAGC 1251  
Qy 129 CCGCAACTGCGCGCCCGCCCGCAGAGAGGCTGTGGAAGTCCGCAAGGAGGCGCCACCA 188  
Db 1252 CCGCAACTGCGAGGCGCCCTAGGAAAAAGGGCTGTGGAATGTGGAAGGAGGACACCA 1311  
Qy 189 GATGAAGACTGCAACCGAGCGCCAGGCAACTCTTCGCGAGGACCTGGGCTTCCGCCCA 248  
Db 1312 AATGAAGATTGTACTGAGAGACAGGCTAA-TTTTTAGGGAAGATCTGGCTTCCCCACA 1370  
Qy 249 GGGCAAGCGCCCGAGTTCCTCCAGCAGAGAGAAACCGCGCCCAACAGCCCAACAGCGCGGA 308  
Db 1371 AGGAAGGCCAGGGAATTTCTTCAGAGCAGACAGAGCCAAACAGCCCAACAGAGAGA 1430  
Qy 309 GCTCAGGTGGCGG-----CGACAACCCCGCAGCGAGCGCGCGCGCGAGCGCCAGGG 362  
Db 1431 GCTTCAGGTTTGGGGAAGAGACAACAACTCCCTCTCAGAGCAGGAGCGGATAGCAAGG 1490  
Qy 363 CA-----CCCTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCTCTGGTGAGCATCAA 416  
Db 1491 AACTGTATCTTTAGCTTCCCTCAGATCACTTTTGGCAGCGACCCCTCTCGTCAANTAAA 1550  
Qy 417 GGTGGCGGCGCAGATCAAGGAGGCCCTGTGACACACCGCGGCCGACGACACCGTGTGGA 476  
Db 1551 GATAGGGGGCAGCTCAAGGAGGCTCTCTGGACACACCGAGCAGACACCGTGTGGA 1610  
Qy 477 GGAGATGAGCTGCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCAT 536  
Db 1611 GGAGATGCTGTGTCAGGCGCGCTGGAAAGCCCAAGATGATCGGCGGGAATTCGGCGGTTTCT 1670  
Qy 537 CAAGTGGCGCAGTACCAACAGATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCAC 596  
Db 1671 CAAGTGGCGCAGTATGACAGATCTCTATCGAATCTGCGGCCCAAGGCTATCGGTAC 1730  
Qy 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCAACATGTGTGACCCAGCTGGG 656  
Db 1731 CGTGTGTTGGGCGCCACACCCCGTCAACATCATCGGACGCAACCTGTTGACCGAGATCGG 1790  
Qy 657 CTGCACCTGAACTTCCCGCATCAGCCCGCATCGACCGTGGCCGTGAAGTGAAGCCCGG 716  
Db 1791 TTGCACGCTGAACCTTCCCGCATTAGCCCTATCGAGACGTTACCCGTTGAAGTGAAGCCCGG 1850  
Qy 717 CATGAGCGGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGAGAGAGATCAAGGCGCTGAC 776  
Db 1851 GATGGACGGCGCCGAGGTCAAGCAATGCGCAATGACAGAGAGAGAGATCAAGGCACTGGT 1910

777 CGCATCTGCGAGGATGAGAGGAGGCAAGATCACCAAGATCGGCCCGCGAAGAACCC 836  
 1911 GGAGATTGACAGAGATGGAAGAGGAGAAATCTCAAGATTGGGCTTGAGAACCC 1970  
 837 CTAACAACACCCCGTGTTCGCCATCAAGAAAGAGAGCAGCACCAAGTGGCGCAAGCTGGT 896  
 1971 GTACAACACCGCGTGTTCGCAATCAAGAAAGAGGACTCGACGAATGGCGCAAGCTGGT 2030  
 897 GGAATTCGGGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGTGGGCAATCCC 956  
 2031 GGAATTCGGGAGCTGAACAGCGCACCGAAGACTTCTGGGAGGTTCAGCTGGGCAATCCC 2090  
 957 CCACCCCGCGCGCTGAAGAAAGAGAGCGTGAACCGTCTGCAAGTGGGCGACGCGCTA 1016  
 2091 GCACCCCGCGGCTGAAGAAAGAGAAATCCGTGACCGTACTGATGTGGGTGATGCCTA 2150  
 1017 CTTACGGTGCCTCGGAGAGGACTTTCGCAAGTACACCGCTTTCACCAATCCCGAGCAT 1076  
 2151 CTTCTCGTTCGCTCGAGAGACTTCAGAAAGTACACTGCTTTCACCAATCCCTTCGAT 2210  
 1077 CAACAGGAGACCCCGGATCGCTACAGTACAAAGTCTGCTGCCCGAGGCTGGAAGG 1136  
 2211 CAACAGGAGACCCCGGATCGATATCAGTACAAAGTCTGCTGCCCGAGGCTGGAAGG 2270  
 1137 CAGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCCCGCAA 1196  
 2271 CTTCCCGCAATCTTCAGAGTAGCATGACCAAAATCCTGGAGCCTTTCGCGAAAAGAA 2330  
 1197 CCGGAGATGCTGATCTACAGTACATGAGACCTGTGAGTGGGAGCGACCTGGAGAT 1256  
 2331 CCGGAGATGCTGATCTACAGTACATGAGTGTGATCGTGGGCTCTGATCTAGAGAT 2390  
 1257 CGGCGAGACCCCGGATCAGAGTCCGCAAGCTGCGCAAGCACTCTGGCTGGGCTTCAC 1316  
 2391 AGGCGAGACCCCGCAACAGATCAGAGTCTGCGCCAGACCTCTTGAGTGGGACTGAC 2450  
 1317 CACCCCGCAAGAGCACCAGAGAGAGCCCTTCTGTGATGGGCTACAGCTGCA 1376  
 2451 CACCCCGCAGAGAGCACCAGAGAGAGCTCTCTCTCTGTGATGGGCTACAGCTGCA 2510  
 1377 CCGGCAAGTGTGACCGTGCAGCCCATCGAGTCTGCGGAGAGAGAGTGTGACCGTGAA 1436  
 2511 CCGTGACAAATGACCGTGCAGCTATCGTGTGCGAGAGAGAGAGTGTGAGCTGTCAA 2570  
 1437 CGATCTCAGAGCTGGTGGGAGCTGAGTGGGCGAGCCAGATCTACCCCGGCTCAA 1496  
 2571 CGATCTCAGAGCTGGTGGGAGTGTGAACTGGGCGAGTCAAGTTTACCCAGGATTA 2630  
 1497 GGTGCGCAGCTGTGCAAGCTGTGCGCGCGCAAGGCGCTGACCGACATCTGTGCCCT 1556  
 2631 GGTGAGGAGCTGTGCANAATCTCTCGCGGAAACCAAGGCACTACAGAGTGTATCCCT 2690  
 1557 GACCGAGGAGCGAGTGTGAGTGTGCGGAGAAACCGCGAGATCTGTGCGGAGCCGTGCA 1616  
 2691 AACCGAGGAGCGAGTGTGAACTGTGCGAGAAACCGAGAGATCTTAAAGGAGCCGTGCA 2750  
 1617 CGGCTGTACTACGACCCAGCAGGAGCTGTGGCGGAGATCCAGAGCGAGGCGCACGA 1676  
 2751 CGGCTGTACTATGACCCCTCCAGGAGCTGTGCGGAGATCCAGAGCGAGGCGCAAGG 2810  
 1677 CAGTGGACCTACAGATCTACAGAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGC 1736  
 2811 CAGTGGACCTATCAGATTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGC 2870  
 1737 CAAGATGCGACCGCCGACACCAACGAGCTGGAAGAGCTGACCGAGGCGGTGCAAGAT 1796  
 2871 CCGGATGAGGAGTGTGCGACACTAACAGCTCAAGCAGCTGACCGAGGCGGTGCAAGAT 2930  
 1797 CGCATGAGAGGATCTGTGATCTGGGCGAGAGCCCGCAAGTCTCGCTGCCATCCAGAA 1856  
 2931 CACCGCAAGGATCTGTGATCTGGGGAAGATCTCTAAGTCTCAAGTGTCCCATCCAGAA 2990  
 1857 GGAGACTTGGAGACCTGTGTGAGCCGACTACTGGCAGGCGCACCTGGATCCCGAGTGGGA 1916

2991 GGAACCTGGAAACCTGGTGGACAGATATTGGCAGGCGCACCTGGATTCTGAGTGGGA 3050  
 1917 GTTCGTGAACACCCCGCCCTTGGTGAAGCTGTGGTACCAAGTGGAGAGAGGCCATCAT 1976  
 3051 GTTCGTGAACACCCCGCCCTTGGTGAAGCTGTGGTACCAAGTGGAGAGAGGCCATCAT 3110  
 1977 CGGCGCGGAGACCTTCTAGTGGAGCGCGCGCAACCGCGAGACCAAGATCGGCAAGG 2036  
 3111 GGGCGCGGAAACCTTCTAGTGGATGGGCGCGTAAACAGGAGACTTAAGTGGGCAAGC 3170  
 2037 CGGCTACGTGAACCGACCGGCGCGGAGAGATCGTGAAGCTTACCGAGACCAACCA 2096  
 3171 CGGATACGTCACTAACCGGCGGAGACAGAAAGTGTTCACCTCACTGACACCAACCA 3230  
 2097 GAAGACCGAGCTGCAGGCCATCCAGTGGCCCTGAGGAGACGCGGAGCGGTGACAT 2156  
 3231 GAAGACTGAGCTGCAGGCCATTTACCTCGCTTTGCAGGACTCGGGCTTGGAGTGAACAT 3290  
 2157 CGTGACCGAGCAGCAGTACGCTGGGCTATCAGAGGCGGCGGCGGCGGAGCGAGAG 2216  
 3291 CGTGACAGACTCTCAGTATGCTTGGGCTATTCAGGCCAGCGAGACCGAGTGAATC 3350  
 2217 CGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAAAGGAGAGTGTACCTGAGCTG 2276  
 3351 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAAAGGAGAGTGTATCTGCGCTG 3410  
 2277 GGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2336  
 3411 GGTACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3470  
 2337 CATCCGCAAGGCTGCTGCTGCGAGCGGCTCGA 2369  
 3471 CATCAGAAAGTGTCTATTCTGATGGCATCGA 3503

RESULT 11

US-09-936-572-12  
 ; Sequence 12, Application US/09936572  
 ; Patent No. 6783981  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UDEN, MARK  
 ; TITLE OF INVENTION: ANTI-VIRAL VECTORS  
 ; FILE REFERENCE: 078883/0137  
 ; CURRENT APPLICATION NUMBER: US/09/936,572  
 ; PRIOR FILING DATE: 2001-12-11  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
 ; PRIOR FILING DATE: 2000-03-17  
 ; PRIOR APPLICATION NUMBER: GB 9906177.2  
 ; PRIOR FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 73  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 12  
 ; LENGTH: 4642  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP2-codon  
 ; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence  
 US-09-936-572-12

Query Match 66.9%; Score 1651.4; DB 4; Length 4642;  
 Best Local Similarity 82.2%; Pred. No. 1.2e-273;  
 Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;  
 12 CATGGCGGAGGCGCATGAGCCAGG---CCACGAGCGCAACATCCTGATGAGCGGAGCAA 68  
 1421 CCTGGCTGAGGCCATGAGCGAGTGACCAACTCCGTACCATCATGATGAGCGGCGCAA 1480  
 69 CTTCAAGGGGCGGCAAGCGGATCATCAAGTGTCTCACTGCGGCAAGAGGCGGCACATCGC 128  
 1481 CTTTCGGAACCAACGCAAGATCGTCAAGTGTCTCAACTGTGGCAAGAGGCGGCACACAGC 1540



Db 3700 GGTACCGCCCAACAAAGGCATTGGCGCAATGAGCAGTGCACAAGCTGCTCGGCTGG 3759  
 QY 2337 CATCCGCAAGGTGCTGTTCTGACGGCATCGA 2369  
 Db 3760 CATCAGAAAGGTGCTATTCTCGATGCGCATCGA 3792

RESULT 12

US-09-552-950-5  
 ; Sequence 5, Application US/09552950  
 ; Patent No. 6541248  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Oxford Biomedica (UK) Limited  
 ; TITLE OF INVENTION: Anti-Viral Vectors  
 ; FILE REFERENCE: 674524-2004  
 ; CURRENT APPLICATION NUMBER: US/09/552,950  
 ; CURRENT FILING DATE: 2000-04-20  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 9772  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:psvNGP  
 US-09-552-950-5

Query Match 66.9%; Score 1651.4; DB 4; Length 9772;  
 Best Local Similarity 82.2%; Pred. No. 1.4e-273;  
 Matches 1951; Conservative 0; Mismatches 406; Indels 16; Gaps 4;

QY 12 CATGGCGGAGCCATGAGCAGG---CCACAGGCGCAACATCCTGATGCGAGCGCAGCAA 68  
 Db 2193 CTTGCTGAGGCCATGAGCAGGTGACCAACTCCGTACCATCATGATGAGCGCGGCAA 2252  
 QY 69 CTTCAAGGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGGAGGCGCACATCGC 128  
 Db 2253 CTTTCGACCAACGCAAGATCGTCAAGTCTTCAACTGTGGCAAGAGGCGCACAGC 2312  
 QY 129 CCACAATGCGGCGCCCCCGCAAGAGGCGTCTGGAAGTGGCGGCAAGGAGGCGCACCA 188  
 Db 2313 CCGCAACTGCGAGGCGCCCTAGGAAAAAGGCGTGTGGAATGTGGAAGAGGAGCACCACCA 2372  
 QY 189 GATGAAGGACTGACCGAGCGCCAGCGCAACTCTTTCGCGAGGACTGCGCTTCCCCCA 248  
 Db 2373 AATGAAGATTGTACTGAGAGACAGGCTAA-TTTTTTTAGGGAAGATCTGGCCCTTCCACA 2431  
 QY 249 GGGCAAGGCGCGGAGTTCCCGAGCAGCAGAAACCGCGCCCAACAGACCCCAACAGCGCGCA 308  
 Db 2432 AGGGAAGGCCAGGGAATTTCTTCAGAGCAGACAGAGCCCAACAGCCCAACAGAGAGA 2491  
 QY 309 GCTGCAAGGTGCGCG-----CGACAAACCCCGCAGCGAGGCGCGCGCGCGAGCGCAGGG 362  
 Db 2492 GCTTCAGGTTTGGGAAGAGACAACTCCCTCTCAGAGCAGGAGCGCGATAGACAAGG 2551  
 QY 363 CA-----CCCTGAATCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGATGATCAA 416  
 Db 2552 AACTGTATCTTTAGTCTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCAATAAA 2611  
 QY 417 GGTGGCGGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCAGACCCGTGTGGA 476  
 Db 2612 GATAGGGGCGACTCAAGAGGCGCTCTCTTGGACAACCGAGCAGACGACCCGTGTGGA 2671  
 QY 477 GGAGATGAGCTCTCCCGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGGCGGTTCAT 536  
 Db 2672 GGAGATGTCGTTGCCAGCGCTGGAAGCCGAAGATGATCGGCGGAATCGGCGGTTCAT 2731  
 QY 537 CAAGTGGCGCAGTACGACCAAGTCTGTGATGAGATCTGGCAAGAGGCGCATCGGCATCGCAC 596  
 Db 2732 CAAGGTGCGCGAGTATGACAGATCTCTCATCGAAATCTGCGGCAACAGGCTATCGGTAC 2791  
 QY 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGGCGCGCAACATGTGACCCAGCTGGG 656

Db 2792 CGTGTGTGGGCGCCACACCCGTCACATCATCGACGCAACCTGTGTGACGAGATCGG 2851  
 QY 657 CTGCACCTTGAACTTCCCATCAGCCCATCAGAGCGGTGCCGTGAAGCTGAGCCCGG 716  
 Db 2852 TTGCACGCTGAACCTTCCCATTAGCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 2911  
 QY 717 CATGGACGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAAGATCAAGGCCCTGAC 776  
 Db 2912 GATGGACGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAAAGATCAAGGCATGGT 2971  
 QY 777 CGCCATCTCGAGGAGATGGAGAGGCAAGATCAACAAATCGGCCCCCGAGAAACC 836  
 Db 2972 GGAGATTTGCACAGAGATGGAAAAGGAAATCTCCAAGATTGGGCTCGAGAACCC 3031  
 QY 837 CTACAAACCCCGTCTTGGCCATCAAGAGAGGACAGACCAAGTGGCGCAAGCTGGT 896  
 Db 3032 GTACAAACAGCGCGGTGTTGCAATCAAGAAAGAGGACTCGACGAAATGGCGCAAGCTGGT 3091  
 QY 897 GGACTTCCGCGAGCTGAAACAAAGCGCACCCAGAGACTTCTGGGAGGTGAGCTGGGCATCCC 956  
 Db 3092 GGACTTCCGCGAGCTGAAACAAAGCGCACCCAGAGACTTCTGGGAGGTGAGCTGGGCATCCC 3151  
 QY 957 CCACCCCGCGGCTCTGAAGAAGAGAGCGCTGACCGTGTGGAAGTGGCGCAAGCGCTA 1016  
 Db 3152 GCACCCCGCAGGCGCTGAAGAAGAAATCCCGTGACCGTACTGATGTGGGTGATGCTA 3211  
 QY 1017 CTTACGCTGCCCCCTGGACGAGGACTTCCGCAAGTACACCGCCTTCAACATCCCAGCAT 1076  
 Db 3212 CTTCTCCGTTCCCTGGACGAAAGCTTTCAGGAAAGTACACTGCTTCAAAATCCCTTCGAT 3271  
 QY 1077 CAAACAGAGACCCCGGCAATCGCTACCAAGTACCAAGTGTCTGCCCGAGGCTGGAGGG 1136  
 Db 3272 CAAACAGAGACCCCGGCAATCGATATCAGTACAGTGTCTGCCCGAGGCTGGAGGG 3331  
 QY 1137 CAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAA 1196  
 Db 3332 CTCTCCGCAATCTTCAGAGTAGCATGACCAAAATCTCGAGGCTTTCGCGCAACAGAA 3391  
 QY 1197 CCGGAGATGCTGATCTACAGTACATGACGACCTGTAGCTGGGCGAGCAGCTGGAGAT 1256  
 Db 3392 CCGGAGATGCTGATCTACAGTACATGATGATGATGATGATGATGATGATGATGATGATGAT 3451  
 QY 1257 CGCCAGCAGCCCGGCAAGATCAGAGAGCTGCGCAAGACCTGCTGCGCTGGGCTTTCAC 1316  
 Db 3452 AGGCGAGCAGCCAGCAGATCAGAGAGCTGCGCAGCAGCTGTTGAGGTGGGAGCTGAC 3511  
 QY 1317 CACCCCGCAACAGAGCAACAGAGAGGAGCCCTTCTCTGTGATGGGCTACAGAGTGCA 1376  
 Db 3512 CACACCGCAACAGAGCAACAGAGAGGAGCCCTTCTCTGTGATGGGCTTACGAGCTGCA 3571  
 QY 1377 CCGGCAAGTGGACCGTGCAGCCCATCGAGCTGCCCGGAGAGAGAGCTGGACCGTGAA 1436  
 Db 3572 CCCTGACAAATGGACCGTGCAGCCTATCGTGTGCGCAGAGAAAGACAGCTGGAGTGTCAA 3631  
 QY 1437 CGACATCCAGAGCTGTGGGCAAGCTGAACCTGGCGCAGCAGATCTTACCCCGGCACTCAA 1496  
 Db 3632 CGACATCAGAGCTGTGGGGAAGTTGAATGGGCGAGTCAAGTATTTACCAGGGATTAA 3691  
 QY 1497 GGTGCGCCAGCTGTGAAGCTGTGCGCGCGCCCAAGGCGCTTCAACGACATCGTGGCCCT 1556  
 Db 3692 GGTGAGGAGCTGTGCAAACTCTCCGCGGAAACCAAGGCACTCACAGAGGTGATCCCGCT 3751  
 QY 1557 GACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCCGAGATCTTGGCGAGCGCGTGCA 1616  
 Db 3752 AACCGAGGAGCGGAGCTCGAACTGGCAGAAAAACCGAGAGATCTTAAAGAGGAGCGGTGCA 3811  
 QY 1617 CGCGGTGTACTACGACCCAGCAAGCACTGTGTGGCGAGATCCAGAAAGCAGGCGCACGA 1676  
 Db 3812 CGCGGTGTACTATGAACCCCTCCAGGACCTGTATCGCGAGATCCAGAGCAGGCGCAAGG 3871  
 QY 1677 CCAGTGGACCTACAGATCTTACAGAGGCGCTTCAAGAACTTCAAGAGCAGGCGCAAGTACGC 1736

Db 3872 CCAGTGGACCTATCAGATTTACAGGAGCCCTTCAAGAACTGAAGACCGCAAGTAGGC 3931  
Qy 1737 CAAGATGGCACCAGCCACCAACGACGCTGAAGAGCTGACCGAGCGCTGCAGAGAT 1796  
Db 3932 CCGATGAGGGGTGCCACACACTAAGCAGCTCAAGCAGCTGACCGAGCGCTGCAGAGAT 3991  
Qy 1797 CGCATGGAGGACATCGTATCTGGGCGCAGAGACCCCAAGTTCCGCTGCCATCCAGAA 1856  
Db 3992 CACCACCGAAAGCATCGTATCTGGGGAAGACTCCCTAAGTTCAAGCTGGCCATCCAGAA 4051  
Qy 1857 GGAGACCTGGGAGACCTTCTAGTGGACCGGACTACTGGCAGGCGCACCTGGATCCCGAGTGGGA 1916  
Db 4052 GGAACTTGGGAACCTTGTGGACAGATATTGGCAGGCGCACCTGGATCTCTAGTGGGA 4111  
Qy 1917 GTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTGATCCAGCTGGAGAGGCCATCAT 1976  
Db 4112 GTTCGTGAACACCCCTCCCTGTGTGAAGCTGTGTGATCCAGCTGGAGAGGCCATCAT 4171  
Qy 1977 CGGCGCGAGACCTTCTAGTGGACCGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2036  
Db 4172 GGGCGCGGAACCTTCTAGTGGATGGGCGCGCTTAACAGGAGACTAAGCTGGGCAAGC 4231  
Qy 2037 CGGTACTGTACGACCGGCGCGGAGAGATCGTGAAGCTGTGACCGAGACCAACCA 2096  
Db 4232 CGGATAGCTACTAACCGGCGGAGAGAGAGTTGTCACTCTGACACCAACCA 4291  
Qy 2097 GAAGACGAGCTGACAGCCATCCAGCTGGCCCTGCAGGACAGCGCGAGGTGAACAT 2156  
Db 4292 GAAGACTGAGCTGACAGCCATTTACCTCGCTTTCAGGAGCTCGGCGCTGGAGGTGAACAT 4351  
Qy 2157 CGTGACGACGACGACGCTGAGTATGCTTGGGCGCATCAATCAAGCCCGAGCGACGAGTGAATC 4411  
Db 4352 CGTGACGACGACGCTGAGTATGCTTGGGCGCATCAATCAAGCCCGAGCGACGAGTGAATC 4471  
Qy 2217 CGAGCTGCTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTG 2276  
Db 4412 CGAGCTGCTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTATCTGCGCTG 4471  
Qy 2277 GGTGCGCGCCCAAGGGCATCGCGGCAACGAGCAGATCGACAGCTGTGTGAGCAAGGG 2336  
Db 4472 GGTACCGCGCCCAAGGGCATTTGGCGCAATGAGCAGCTCGACAAAGCTGTGCTCGGCTG 4531  
Qy 2337 CATCCGCAAGGTGCTGTTCTGGACGCGCATCGA 2369  
Db 4532 CATCAGGAAGGTGCTATTTCTGGATGTCATCGA 4564

## RESULT 13

US-09-872-733A-6  
; Sequence 6, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; TITLE OF INVENTION: SIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872.733A  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 8366  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
; OTHER INFORMATION: of the construct pCMVgagpolenkan containing a CMV  
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin  
; OTHER INFORMATION: resistance gene

## US-09-872-733A-6

Query Match 64.5%; Score 1592.8; DB 4; Length 8366;  
Best Local Similarity 80.7%; Pred. No. 1.4e-263;  
Matches 1914; Conservative 0; Mismatches 442; Indels 16; Gaps 4;  
Qy 14 TGGCGGAGGCGCATGAGCCAGCCACACGAGC---GCCAACATCTCTGATGCGAGCGCAGCACT 70  
Db 1857 TGGCGGAGGCGCATGAGCCAGGTCAGAACTCGGCGACCATTAATGATGCGAGAGGCACT 1916  
Qy 71 TCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGCCACATCGGCC 130  
Db 1917 TCCGGAACACGCGGAAGATCGTCAAGTCTTCAATTTGGCAAGAGGGGACACCGGCA 1976  
Qy 131 GCAATCTCCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCAGGAGAGGGCCACACAGA 190  
Db 1977 GGAATCTCCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGTGGAAGAGGAGACACAAA 2036  
Qy 191 TGAAGGACTGCACCGAGCGCCGACCACTTCTTCCGGAGGACTTGGCCTTCCCGCAGG 250  
Db 2037 TGAAGGATTTGTACTGAGAGACAGGCTAA-TTTTATTAGGGAAGATCTGGCCTTCTTACAG 2095  
Qy 251 GCAAGGCGCGGAGTTCCCGAGCGAGCAGAACCGCCCAACAGCCCAACAGCGCGCAGC 310  
Db 2096 GGAAGGCGCAGGGAATTTCTTTCAGAGCAGACAGCCACAGCCCGCCACAGAGAGAGC 2155  
Qy 311 TGCAGGTGCGCGG-----CGACAAACCCCGAGCGAGCGCGCGCGAGCGCCAGGGCA 364  
Db 2156 TTCAGGTCTGGGCTAGAGACAACTCTCCCTCAGAAAGCAGAGAGCGCATAGACAAAGAA 2215  
Qy 365 -----CCCTGNACTTCCCGCAGATCACCTGTGGCAGCGCCCTTGTGAGCATCAAGG 418  
Db 2216 CTGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCAAGTAAGGA 2275  
Qy 419 TGGCGCGCAGATCAAGGAGCGCTCTGTGACACCGCGCGCAGCAGACACCGTGTGGAGG 478  
Db 2276 TCGGGGGGCACTCAAGAGAGCGCTGTCTGATACAGAGCAGATGATACAGTATTAGAG 2335  
Qy 479 AGATGAGCCTTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGGCGATCGCGGCTTCATCA 538  
Db 2336 AATGAGTTTCCAGGAAGATGGAAACCAAAATATGATAGGGGGATCGGGGCTTCATCA 2395  
Qy 539 AGTGCGCGCAGTACGACAGATCTGATCGAGATCTCGGCGCAAGAGCCCATCGCACCG 598  
Db 2396 AGTGAGGCGAGTACGACAGATCTGATAGAAATCTGTGGACATATAAGTATAGGTACAG 2455  
Qy 599 TGCTGATCGGCGCCCGCCCGTGAACATCATCGGCGCCCAACATCTGACCCAGCTGGGCT 658  
Db 2456 TATTAGTAGGACCTACACCTGTCAACATATTGGAAGAAATCTGTTGACCCAGATCGGCT 2515  
Qy 659 GCACCTCTGAATTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCA 718  
Db 2516 GCACCTTGAATTTCCCATCAGCCCTATTGAGACGGTCCCGTGAAGTTGAAGCCGGGA 2575  
Qy 719 TGAACGCGCCCAAGGTGAAGCAGTGGCCCTTGACCGGAGGAGAAATCAAGGCCCTGACCG 778  
Db 2576 TGAACGCGCCCAAGGTCAAGCAATGGCCATTTGACGAAAGAGAAATCAAGGCCCTTAGTCG 2635  
Qy 779 CCATCTCGGAGGAGATGGAAGAGGAGGCAAGATCACCAAGATCGGCGCCCGAGAACCCCT 838  
Db 2636 AATCTGTACAGAGATGGAAGAGGAGGAGATCAGCAAGATCGGCGCTGAGAACCCCT 2695  
Qy 839 ACAACACCCCGCTTTCGCCATCAAGAGAGAGGAGCAGACCAAGTGGCGCAGCTGGTGG 898  
Db 2696 ACAACACTCCAGTCTTCGCAATCAAGAGAGGAGGAGCAGTACCAGTGGAGAAAGCTGGTGG 2755  
Qy 899 ACTTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTCAGCTGGGCGATCCCC 958  
Db 2756 ACTTCAGAGAGCTGAACAGAGAACTCAGGACTTCTGGGAAGTTTCAGCTGGGCGATCCCAC 2815  
Qy 959 ACCCGCGCGCGCTGAAGAGAGAGAGCGTGCAGCGCTGGAGCTGGCGGCGACCGCTACT 1018  
Db 2816 ATCCCGCTGGGTTGAAGAGAGAGAGTCAAGTACAGCTGCTGGATGTGGGTGATGCCTACT 2875



```
QY 1019 TCAGCGTGCCTTGGACGAGCTTCGCAAGTACACCGCTTCCATCCCCAGCATCA 1078
Db 2876 TCTCCGTTCCCTTGGACGAGACTTCAGGAAGTACACTGCTTTCCAGATACCTAGCATCA 2935
QY 1079 ACAACAGACACCCCGGCATCCGCTACAGTACAACTGTCTGCCCGCAGGGCTGGAAGGCA 1138
Db 2936 ACNACAGACACACAGGATCCGCTACAGTACAACTGTCTGCCCGCAGGGATGGAAGGAT 2995
QY 1139 GCCCCAGCATTTTCCAGAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCGCCCGCAACC 1198
Db 2996 CACCAGCCATCTTTCAAAGCAGCATGACCAAGATCTCTGAGCCCTTCCGCAAGCAAAACC 3055
QY 1199 CCGAGATCTGTATCTACAGTACAGCAGACTGTAGCTGGCGAGGACCTTGGAGATCG 1258
Db 3056 CAGACATCTGTATCTATCATAGTACAGCAGACTCTACGTAGGAAGTGAACCTGGAGATCG 3115
QY 1259 GCAGCACCAGCGCAAGATCTGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGGCTTCACCA 1318
Db 3116 GGCAGCACAGGACCAAGATCTGAGGAGCTGAGCAGCATCTGTTGAGGTGGGACTGACCA 3175
QY 1319 CCCCCAAGAAAGACCAAGAGGAGCCCCCTTCTCTGTGGATGGGCTAGAGCTGACC 1378
Db 3176 CACCAGACAAGAGCAACCAAGGAACTCTCCCTTCTGTGTGGATGGGCTACGAATCTGCATC 3235
QY 1379 CCGAAGTGGACCTGCGAGCCCATCGAGCTGCCCGAGGAGAGCTGGACCGTGAACG 1438
Db 3236 CTGACAAGTGGACAGTGGAGCCCATCGTCTGCTGAGAGGAGACGTGAGACTGTGAACG 3295
QY 1439 ACATCCAGAAGCTGTGTGGGCAAGCTGAATCTGGGCCAGCCAGATCTTACCCGGCATCAAG 1498
Db 3296 ACATACAGAAGCTGTGTGGGCAAGTGAATCTGGGCCAAGCCAGATCTTACCCAGGATCAAG 3355
QY 1499 TGGCCAGCTGTCAAGCTGTCTGCGCGCGCCCAAGGCCCTTGACCGCATCTGTGCCCTGA 1558
Db 3356 TTAGGCAGCTGTCAAGCTGTCTGAGGAACCAAGGCACTTGACAGAAGTGAATCCCACTGA 3415
QY 1559 CCGAGAGCCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGCCGCTGCAG 1618
Db 3416 CAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCTCTGAAGGAGCCAGTACATG 3475
QY 1619 GCCTGTACTACGACCCAGCAAGGACCTGTGTGCGCGAGATCTCCAGAAGCAGGGCCAGACC 1678
Db 3476 GAGTGTACTACGACCCAGCAAGGACCTGTATCGCAGAGATCCAGAGCAGGGGCCAAGGCC 3535
QY 1679 AGTGGACCTACCAAGTCTTACAGGAGCCCTTTAAGAACCTTGAAGACCGGCAAGTACGCCA 1738
Db 3536 AATGGACCTTACCAAACTTACCAAGGAGCCCTTTCAAGAACCTTGAAGACAGGCAAGTACGCCA 3595
QY 1739 AGATGCGCACCGCCACACCAAGCAGTGAAGCAGCTGACCGAGGCGGTGCGAGAAGATCG 1798
Db 3596 GGATGAGGGGTGCCCCACACCAACATGTGAAGCAGCTGACAGAGCAGTGCAGAAATCA 3655
QY 1799 CCATGAGAGCAGCATCTGTATCTGGGGCAAGACCCCAAGTTTCCGCTGCCCATCCAGAAGG 1858
Db 3656 CCACAGAGAGCATCTGTATCTGGGGCAAGACTTCCCAAGTTTCAAGCTGCCCATACAGAGG 3715
QY 1859 AGACTTGGAGACCTGTGTGACCGCATCTA CTGGCAGGCGCACCTGTGATCTCCCGAGTGGAGT 1918
Db 3716 AGACATGGGAGACATGTTGTGACCGAGTACTTGGCAAGCCACTTGGATCTCTGTGAGTGGAGT 3775
QY 1919 TCGTGAACACCCCTTGTGAAGCTGTGTGTTACAGCTGGAAGAGGAGGCCCATCATCG 1978
Db 3776 TCGTGAACACCCCTTGTGTGAAGCTGTGTGTTACAGCTGGAAGAGGAGGCCCATCTGTGG 3835
QY 1979 GCGCCGAGACCTTTACTGTGGAAGCGGGCCGCCCAACCGCAGACCAAGATCGGAAGGCCG 2038
Db 3836 GAGCAGAGACCTTTACTGTGATGTTGGGCGAGCCAAACAGGGAGACCAAGCTGGGCAAGGCAG 3895
QY 2039 GCTACTGTGACCGACCGGGCGCGCAGAGATCTGTGAGCTGTGACCGCAGACCAACCAACGAG 2098
Db 3896 GCTACTGTGACCAACCGGAGGACGACAGAAAGTGTGTGACCTGTGACTGACACCAACCAACGAG 3955
```

```
QY 2099 AGACGAGCTGCAGGCCATCCAGCTGGCCCTGCAGACAGCGCAGAGGTGAACATCG 2158
Db 3956 AGACTGAGCTGCAGGCCATCTACCTAGCTTCTGCAAGACAGCGACTGGAAGTGAACATCG 4015
QY 2159 TGACCGACAGCCAGTACGCTTGGGCATCATCCAGGCCCGCAGCCGACAGAGCGAGAGCG 2218
Db 4016 TGACGACTCACAGTACGCACTGGGCATCATCCAGCACAAACAGACCAATCCGAGTCAAG 4075
QY 2219 AGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGG 2278
Db 4076 AGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTGTGGCATGGG 4135
QY 2279 TGCCCCCCACAGAGGCGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGCAGCAAGGGCA 2338
Db 4136 TACCAGCACAAAGAAATTGGAGGAATGAACAGTAGATAAATTAGTCAGTGTCTGGGA 4195
QY 2339 TCCGCAAGGTGCTGTTCTCTGGAGCGGCATCGAT 2370
Db 4196 TCCGGAAGGTGCTGTTCTCTGGAGCGGATCGAT 4227

RESULT 14
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

Query Match 63.1%; Score 1557.4; DB 4; Length 4338;
Best Local Similarity 80.5%; Pred. No. 1.5e-257;
Matches 1914; Conservative 0; Mismatches 441; Indels 22; Gaps 7;

QY 14 TGGCCGAGGCGCATGAGCCAGGCCACCAGC---GCCAACATCTCTGATGTCAGCGCAGCAACT 70
Db 1085 TGGCCGAGGCGCATGAGCCAGGTCAGCACTCGCGCGACCAATAATGATGTCAGAGAGGCAACT 1144
QY 71 TCAAGGGCCCCCAAGCGCATCATCAAGTGTCTTAACTTCGGCAAGGAGGGGCCAATCGCCCC 130
Db 1145 TCCGGAACACAGCGGAAGATCGTCAAGTGTCTTAATTGTGGCAAGAAGAGGGCAACCCGCCA 1204
QY 131 GCAACTGCCCGCGCCCCCGCAAGAGGGCTGTGTGAAGTGTGCGCAAGGAGGGCCACCAAGA 190
Db 1205 GGAACCTGCCGGGGCCCCCGGAAGAGGGCTGTTTGGAAATGTGGAAAGGAAGGACACCAAA 1264
QY 191 TGAAGACTGTCACCGAGCGCCAGGCCAACTTCTTCCGCGAGGACCTGGCCCTTCCCCCAGG 250
Db 1265 TGAAGATTTGTACTGAGAGACAGGCTTAA-TTTTATTAGGGGAAGATCTGGCCCTTCTCAAG 1323
QY 251 GCAAGGCCCGCGAGTTTCCCCAGCGAGCAGAAACCGCGCAACAGCAGCCCAACAGCGCGAGC 310
Db 1324 GGAAGGCCAGGGAAATTTTCTTCAAGAGCAGACACAGAGCCACAGCCCAACCAAGAGAGAGC 1383
QY 311 TGCAGGTGCGCGG-----CGAACACCCCGCAGCAGAGGCCGCGCGCAGCGCCAGGGCA 364
```

Db 1384 TTCAGGTCGGGTAGACACAACTCCCTCAGAAAGCAGGAGCCGATAGACAAAGGAA 1443  
Qy 365 -----CCCTGAACCTCCCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGG 418  
Db 1444 CTGTATCCTTTAACTTCCCTCAGATCACTTTGGCAACGACCCCTCGTACAGTAAGGA 1503  
Qy 419 TGGGGGCGCAGATCAAGAGGCGCTGTGGACACCGGCGCGACACACCGTGTGTGGAGG 478  
Db 1504 TCGGGGGGCAACTCAAGGAAGCGTGTCTGATACAGGAGCAGATGATACAGTATTAGAA 1563  
Qy 479 AGATGAGCGCTCCCGGAAGTGAAGCCAAAGATGATCGCGGCGATCGCGGCTTCATCA 538  
Db 1564 AAATGAGTTTCCAGGAAGATGGAACCAAAATGATAGGGGGATCGGGGGCTTCATCA 1623  
Qy 539 AGGTGCGCCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACG 598  
Db 1624 AGGTGAGGACGTACGACAGATCTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683  
Qy 599 TGCTGATCGGCCCCACC-----CCGTGAACATCATCGGCCCAACATGCTGACCCAGCTG 654  
Db 1684 TATTAGTAGGACCTACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATC 1743  
Qy 655 GGCTGCACCTGAACCTTCCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCC 714  
Db 1744 GGCTGCACCTTGAACCTTCCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCG 1803  
Qy 715 GGCATGAGCGGCCCAAGGTGAAGCAGTGGGCCCTTGACCGAGGAGAAATCAAGGCCCTG 774  
Db 1804 GGGATGAGCGGCCCAAGGTCAAGCAATGGCCATTGACAAAGAGAGATCAAGGCCCTTA 1863  
Qy 775 ACCGCGATCTCGAGGAGATGAGAGAGGGGGAAGATCACAAAGATCGGCCCGAGAAC 834  
Db 1864 GTCGAAATCTGTACAGAGATGAGAGAGGAGGAGAGATCAGAAAGATCGGGCCCTGAGAAC 1923  
Qy 835 CCCTTACAAACCCCGCTGTCCCATCAAGAGAGGAGACACCAAGTGGCGCAAGCTG 894  
Db 1924 CCCTTACAAACTCTCAGTCTTCCGAATCAAGAGAGGAGACATACCAAGTGGAGAAAGCTG 1983  
Qy 895 GTGACATCTCCGAGCTGAACAAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954  
Db 1984 GTGACATCTCAGAGCTGAACAAAGAGACTCAGGACTTCTGGGAAGTTACGCTGGGCATC 2043  
Qy 955 CCCAACCCCGCGCTGAAGAGAGAGAGAGCTGACCGTGTGGAGCTGGGCGAGCGCC 1014  
Db 2044 CCACATCCCGCTGGGTGAAGAGAGAGAGTCAGTGACAGTGTGATGTGGGTGATGCC 2103  
Qy 1015 TACTTCAGCTGCGCTCGAGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGC 1074  
Db 2104 TACTTCTCCGTTGGACGAGACTTTCAGGAAGTACATGCGCTTCCAGATACCTAGC 2163  
Qy 1075 ATCAACACGAGACCCCGGCTACCGCTACCAAGTACACCGTGTGCGCCCGAGGCTGGAAG 1134  
Db 2164 ATCAACACGAGACCCAGGCTACCGCTACCAAGTACCAAGTGTGCGCCAGGATGGAAG 2223  
Qy 1135 GGCAGCCCCAGATCTTTCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCGCGCCGCG 1194  
Db 2224 GGATCACCGCCATCTTTCAGAGCAGCATGACCAAGATCTGAGAGCCCTTCCGCAAGCAA 2283  
Qy 1195 AACCCGAGATCTGTATCTACAGTACATGAGCAGCCTGTACGTTGGGCGAGCCTGGAG 1254  
Db 2284 AACCCGAGATCTGTATCTACAGTACATGAGCAGCCTGTACGTTGGGAGTGAATCTGGAG 2343  
Qy 1255 ATC-GGGCAGCACCGCGCAAGATCTGAGGAGCTGCGCAAGCCTGTGCGCTGGGGCTT 1313  
Db 2344 ATCGGGGCGAGCACAGGACCAAGATCTGAGGAGCTGAGACAGCATCTGTTGAGGTGGGACT 2403  
Qy 1314 CACCACCCCGACAGAGACCAAGAGAGGCCCTTCTGTGGATGGGCTACGAGCT 1373  
Db 2404 GACCACACCAAGAGACCAAGAGAGACCTCCCTTCTGTGGATGGGCTACGAACT 2463  
Qy 1374 GCACCCCGAAGTGAAGCTGCGAGCCCATCGAGCTGCCCGAGAGGAGCTGGACCGT 1433  
Db 2464 GCATCTTGAAGTGGACAGTGGAGCCCATCTGCTGCTGAGAGGAGACAGCTGGAGCTGT 2523

Qy 1434 GAACGACATCCAGAAAGCTGTGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCCGCAT 1493  
Db 2524 GAACGACATCAGAAAGCTCGTGGGCAAGTGAATCTGGGCAAGCCAGATCTTACCCAGCAT 2583  
Qy 1494 CAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCCAAAGCCCTGAGCCGACATCGTGCC 1553  
Db 2584 CAAGGTAGGCGCTGTGCAAGCTGTGCGAGGAACTGAGGCACTGACAGAAAGTGTATCCC 2643  
Qy 1554 CCTCAGCAGAGAGCCCGAGCTGTGAGCTGGCCGAGAACCCGCGAGATCTCTGCGCGAGCCCGT 1613  
Db 2644 ACTCAGAGAGGAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCTCTGAAGAGCCAGT 2703  
Qy 1614 GCACGGGTGTACTACGACCCCAAGAGAACCTGGTGGCCGAGATCTCAAGAGCAGGGCCA 1673  
Db 2704 ACATGGAGTGTACTACGACCCCAAGAGAACCTGATCGCAGAGATCCAGAAAGCAGGGCA 2763  
Qy 1674 CGACAGTGGACCTTACAGATCTTACAGGAGCCCTTCAAGAACCTGAGAGCCGCAAGTA 1733  
Db 2764 AGGCAATGGACCTTACCAATCTTACAGGAGCCCTTCAAGAACCTGAGAGCAGGCAAGTA 2823  
Qy 1734 CGCCAGATGCGCACCCCAACACAGCTGAAGCAGCTGACCGAGGCGCTGCAGAA 1793  
Db 2824 CGCAAGATGAGGGTGGCCACACACAGATGTGAGCAGCTGACAGGCGAGTGCAGAA 2883  
Qy 1794 GATCGCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTTGCCTCCATCCA 1853  
Db 2884 GATCACACAGAGAGCATCTGTGATCTGGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943  
Qy 1854 GAAGGACCTGGGAGACCTGTGTGACCGGCTGTGAGCAGCTGACCGAGGCGCTGGAGTGG 1913  
Db 2944 GAAGGACATGAGGAGACATGTGTGACCGGCTGTGAGCAGCTGACCGAGGCGCTGGAGTGG 3003  
Qy 1914 GGAGTTCTGTAACACCCCTTGTGTGAGCTGTGTGACCGCTGTGAGAGAGGAGCCCAT 1973  
Db 3004 GGAGTTCTGTAACACCCCTTGTGTGAGCTGTGTGAGCTGTGTGAGAGAGGAGCCCAT 3063  
Qy 1974 CATCGCGCGCAGACCTTCTAAGTGTGAGCGCGCGCCCAACCGCAGAGCAAGATCGGCAG 2033  
Db 3064 CGTGGAGCAGAGACCTTCTACGTGGATGGGGCAGCCAAACAGGGAGACCAAGCTGGGCAA 3123  
Qy 2034 GGCAGGCTGTGACACCGGCGCGGCGCAGAGATCTGTGAGCTGTGAGCAGGAGCCCA 2093  
Db 3124 GGCAGGCTGTGACACCGGCGCGGCGCAGAGATCTGTGAGCTGTGAGCAGGAGCCCA 3183  
Qy 2094 CCAGAGACCGAGCTGCAGGCTTCCAGCTGGCGCTTCAGGACAGCGCGCAGCGAGGTGAA 2153  
Db 3184 CCAGAGACTGTGAGCTGCAGGCTTCCAGCTGTGAGCTGTGAGAGTGAAGTGA 3243  
Qy 2154 CATCGTACCGACAGCGAGTACGCTTGGGCTATCATCGCGCCATCATCGCGCCAGCCGCA 2213  
Db 3244 CATCGTACAGACTTCACAGTAGC-CATGGGCTATCATCCAGCACAACAGACCAATCGA 3302  
Qy 2214 GAGGAGCTGTGAGCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAG 2273  
Db 3303 GTCAGAGCTGTGAGCAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGGC 3362  
Qy 2274 CTGGGTCCCGCCCAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGTAGCAA 2333  
Db 3363 ATGGGTACAGACACAAAGGATTTGAGGAGATGAAACAGTAGATTAATTTAGTCAGTGC 3422  
Qy 2334 GGGCATCCGAAAGTGTCTTCTTGGAGCGGCATCGAT 2370  
Db 3423 TGGGATCCGAAAGTGTCTTCTTGGAGCGGCATCGAT 3459

RESULT 15  
US-09-952-060-1  
; Sequence 1, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emili, Emilio A.  
; APPLICANT: Youil, Rima

; APPLICANT: Bett, Andrew J.  
; APPLICANT: Chen, Ling  
; APPLICANT: Kaslow, David C.  
; APPLICANT: Shiver, John W.  
; APPLICANT: Toner, Timothy J.  
; APPLICANT: Casimiro, Danilo R.  
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
; TITLE OF INVENTION: MODIFICATIONS  
; FILE REFERENCE: 20747Y  
; CURRENT APPLICATION NUMBER: US/09/952,060  
; CURRENT FILING DATE: 2001-09-14  
; PCT APPLICATION NUMBER: PCT/US01/28861  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/317,814  
; PRIOR FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: 60/279,056  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/233,180  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2577  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Codon optimized DNA encoding modified wt pol  
US-09-952-060-1

Query Match 54.1%; Score 1336.4; DB 4; Length 2577;  
Best Local Similarity 86.7%; Pred. No. 8.3e-220;  
Matches 1472; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

Qy	672	CCCCATCAGCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCGCGCATGAGCGGCCCAA	731
Db	15	CCCCATCTCCCCATTGAGACTGTGCCTGTGAAGCTGAAGCCCTGGCATGGATGGCCCAA	74
Qy	732	GGTGAAGCAGTGGCCCTGACCGAGGAGATCAAGGCCCTCACGCCATCTGCGAGGA	791
Db	75	GGTGAAGCAGTGGCCCTGACTGAGGAGAGATCAAGGCCCTGGTGAATCTGCACTGA	134
Qy	792	GATGGAGAGAGGCGCAAGATCAACAGATCGGCCCGCGAACCCTTACAACACCCCGCT	851
Db	135	GATGGAGAGAGGCGCAAAATCTCAAGATTGGCCCCCGAGAACCCCTTACAACACCCCTGT	194
Qy	852	GTTCCGCATCAAGAGAGACAGACCAAGTGGCGGCAAGCTGGTGGACTTCGCGAGCT	911
Db	195	GTTTGCCATCAAGAAGAGACTCCACCAAGTGGGGAAGCTGGTGGACTTCAAGGAGCT	254
Qy	912	GAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCGCCCT	971
Db	255	GAACAAGAGACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCTGGCCT	314
Qy	972	GAAGAAGAAAGAGCGTGACCGTGTGAGCGTGGCGAGCGCCCTACTTCAAGCGTGCCT	1031
Db	315	GAAGAAGAAAGTCTGTGACTGTCTGGATGTGGGGAGTGCCTACTTCTGTGCCCCCT	374
Qy	1032	GGACAGGACTTCGCGAAGTACACCGCTTTCACATCCCCAGCATCAACAAAGAGACCCC	1091
Db	375	GGATGAGGACTTCAGGAAGTACACTGCTCTTTCACCATCCCCCTCCATCAACAAATGAGACCCC	434
Qy	1092	CGGCATCCGTACAGTACAACTGTGCCCCAGGCTGGAAGGGCGAGGCCAGCATCTT	1151
Db	435	TGGCATCAGGTACAGTACAACTGTGCCCCAGGCTGGAAGGGCTTCCCCCTGCCATCTT	494
Qy	1152	CCAGAGCAGCATCAACAAAGTCTTGGAGCCCTTTCGCGCCCGCAACCCCCAGATCTGTAT	1211
Db	495	CCAGTCTCTATGACCAAGATCTTGGAGCCCTTTCAGGAAGCAGAACCTTGACATTTGTAT	554
Qy	1212	CTACCAAGTACATGGAACGACTGTGAGTGGGACGAGCCTGGAGATCGGCACGACCGCG	1271
Db	555	CTACCAAGTACATGGAACGACTGTGAGTGGGCTCTGACCTGGAGATTGGGCGAGCACAGGAC	614

Qy	1272	CAAGATCGAGGAGCTGCGCAAGCACCTGTGGCTGGGGCTTTCACCAACCCCGCACAGAA	1331
Db	615	CAAGATTGAGGAGCTGAGGCGAGCACCTGTGAGTGGGGCTGACCAACCCCTGACAAGAA	674
Qy	1332	GCACAGAAGGAGCCCCCTTCTGTGATGGGTACGAGCTGCACCCCGACAAGTGGAC	1391
Db	675	GCACAGAAGGAGCCCCCTTCTGTGATGGGTATGAGCTGCACCCCGACAAGTGGAC	734
Qy	1392	CGTGACGCCATCGAGCTGCCGAGAGGAGAGCTGAGACCGTGAACGACATCCAGAGCT	1451
Db	735	TGTGAGCCCATTTGTCTGCTGAGAAAGGACTCTCTGACTGTGAATGACATCCAGAGCT	794
Qy	1452	GGTGGCAAGCTGAACTGGGCCAGCCAGATCTACCCCGGCATCAAGGTGCGCAGCTGTG	1511
Db	795	GGTGGCAAGCTGAACTGGGCCCTCCCAATCTACCTTGGCATCAAGGTGAGGCGAGCTGTG	854
Qy	1512	CAAGCTGTGCGCGCGCAAGCCCTGACCGACATCTGTGCTCCCTGACCGAGAGGCGGA	1571
Db	855	CAAGCTGTGAGGGGCAACCAAGGCCCTGACTGAGGTGATCCCCCTGACTGAGGAGGCTGA	914
Qy	1572	GCTGAGCTGGCCGAGAACCGCGAGATCTGCGGAGCCCGTGACGGCGTGTACTACGA	1631
Db	915	GCTGAGCTGGCTGAGAACAGGAGATCTTGAAGAGGCCCTGTGCTGCGGTGTACTATGA	974
Qy	1632	CCCCAGCAAGGACCTGTGGCCGAGATCCAGAAGCAGGGGCCAGCACAGTGGACCTACCA	1691
Db	975	CCCCCTCAAGGACCTGATTTGCTGAGATCCAGAAGCAGGGGCCAGGGCCAGTGGACTACCA	1034
Qy	1692	GATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGC	1751
Db	1035	AATCTACAGGAGCCCTTCAAGAACCTGAAGACTGGCAAGTATGCCAGGATGAGGGGGC	1094
Qy	1752	CCACACCAAGCAGCTGAACGACGTGACCGGCGCGTGCGAGAGATCGCCATGAGAGCAT	1811
Db	1095	CCACACCAATGATGTGAAGCAGCTGAGGTGTGCGAAGATCACCACCTGAGTCCAT	1154
Qy	1812	CGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCTCCATCCAGAAGGAGACCTGGGAGAC	1871
Db	1155	TGTGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCCTCCATCCAGAAGGAGACCTGGGAGAC	1214
Qy	1872	CTGCTGGAACGACTACTGCGAGGCCACCTGGATCTCCCGAGTGGGAGTTTGTGAACACCC	1931
Db	1215	CTGCTGGAACGACTACTGCGAGGCCACCTGGATCTCCCGAGTGGGAGTTTGTGAACACCC	1274
Qy	1932	CCCCCTGGTGAAGCTGTGTTACAGCTGGAGAGGAGCCCATCATCGGSCCGAGACCTT	1991
Db	1275	CCCCCTGGTGAAGCTGTGTTACAGCTGGAGAGGAGCCCATTTGTGGGGCTGAGACCTT	1334
Qy	1992	CTACGTGGAGCGGCGCCCAACCGCGAGACCAAGATCGGCAAGGCCCGGCTACGTGACCGA	2051
Db	1335	CTATGTGATGGGCTGCCAACAGGGAGACCAAGCTGGGCAAGGCTGGCTATGTGACCAA	1394
Qy	2052	CGGGGCGCGGAGAAATCGTGAGCCTGACCGAGACCAACCAAGAGCCAGAGCTGCA	2111
Db	1395	CAGGGCGGCGGAGAAATCGTGAGCCTGACCTGACCTGACCAACCAAGAGCTGCA	1454
Qy	2112	GGCCATCCAGCTGGGCCCTGCGAGACGCGGAGGTGAACATCGTGACCCGACAGCCA	2171
Db	1455	GGCCATCTACCTGGGCCCTCAGGACTCTTGGCTTGGAGGTGAACATTTGTGACTGCTCCCA	1514
Qy	2172	GTACGCCCTTGGGCTCATCATCAGGCCCGAGCCCGCAAGAGAGCGAGCGAGCTGGTGAACCA	2231
Db	1515	GTATGCCCTTGGGCTCATCATCAGGCCCGAGCTGATCATCTGAGTCTGAGCTGGTGAACCA	1574
Qy	2232	GATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCGGCCCAACA	2291
Db	1575	GATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCGGCCCAACA	1634
Qy	2292	GGGCATCGGGGCAAGCAGCATCCACAAGCTGTGAGCAAGGCGATCCGCAAGGTGCT	2351
Db	1635	GGGCATCGGGGCAATGAGCAGGTGCAAGCTGGTGTCTGCTGGCATCAGGAAGGTGCT	1694

Qy 2352 GTTCCTGGACGGCATCGA 2369  
|||||||  
Db 1695 GTTCCTGGATGGCATTGA 1712

Search completed: June 2, 2005, 03:12:53  
Job time : 401.948 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 09:47:50 ; Search time 389 Seconds  
(without alignments)  
10360.282 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463  
Sequence: 1 gtgcagccaccatggcgca.....gggtagcagcggtgaattc 2463

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2046	83.1	2306	4	US-09-475-515-82 Sequence 82, Appl
2	2025.2	82.2	2312	4	US-09-475-515-84 Sequence 84, Appl
3	2019.2	82.0	2300	4	US-09-475-515-83 Sequence 83, Appl
4	1942.6	78.9	4319	4	US-09-475-515-6 Sequence 6, Appl
5	1878.8	76.3	2305	4	US-09-475-515-80 Sequence 80, Appl
6	1852	75.2	2299	4	US-09-475-515-81 Sequence 81, Appl
7	1651.8	67.1	4307	4	US-09-552-950-2 Sequence 2, Appl
8	1632.6	66.3	4307	4	US-09-552-950-2 Sequence 2, Appl
9	1624.6	66.0	4327	4	US-09-936-572-14 Sequence 14, Appl
10	1624.6	66.0	4353	4	US-09-936-572-13 Sequence 13, Appl
11	1624.6	66.0	4642	4	US-09-936-572-12 Sequence 12, Appl
12	1624.6	66.0	9772	4	US-09-552-950-5 Sequence 5, Appl
13	1566	63.6	8366	4	US-09-872-733A-6 Sequence 6, Appl
14	1530.6	62.1	4338	4	US-09-872-733A-1 Sequence 1, Appl
15	1309.6	53.2	2577	4	US-09-952-060-1 Sequence 1, Appl
16	1309.6	53.2	2650	4	US-09-952-060-5 Sequence 5, Appl
17	1307	53.1	4053	4	US-09-952-060-34 Sequence 34, Appl
18	1304.8	53.0	2577	4	US-09-952-060-3 Sequence 3, Appl
19	1304.8	53.0	2650	4	US-09-952-060-7 Sequence 7, Appl
20	1304.8	53.0	38519	4	US-09-952-060-28 Sequence 28, Appl
21	1203.6	48.9	9010	4	US-09-184-418C-8 Sequence 8, Appl
22	1175.8	47.7	9913	4	US-09-827-688-11 Sequence 11, Appl
23	1172.6	47.6	8972	4	US-09-184-418C-9 Sequence 9, Appl
24	1165.4	47.3	8959	4	US-09-184-418C-11 Sequence 11, Appl
25	1163.8	47.3	2467	4	US-09-872-733A-3 Sequence 3, Appl
26	1142	46.4	8992	4	US-09-184-418C-4 Sequence 4, Appl
27	1106.8	44.9	2601	3	US-09-117-217-7 Sequence 7, Appl

28 1106.8 44.9 2601 3 US-09-117-217-9 Sequence 9, Appl  
29 1106.8 44.9 2601 3 US-09-117-217-11 Sequence 11, Appl  
30 1106.8 44.9 2601 3 US-09-117-217-13 Sequence 13, Appl  
31 1106.8 44.9 2601 4 US-09-735-487-7 Sequence 7, Appl  
32 1106.8 44.9 2601 4 US-09-735-487-9 Sequence 9, Appl  
33 1106.8 44.9 2601 4 US-09-735-487-11 Sequence 11, Appl  
34 1106.8 44.9 2601 4 US-09-735-487-13 Sequence 13, Appl  
35 1106.8 44.9 4307 4 US-09-552-950-1 Sequence 1, Appl  
36 1106.8 44.9 4307 4 US-09-936-572-1 Sequence 1, Appl  
37 1106.8 44.9 9719 4 US-09-700-304-1 Sequence 1, Appl  
38 1105.2 44.9 9050 4 US-09-184-418C-7 Sequence 7, Appl  
39 1100.4 44.7 7399 2 US-08-418-848A-9 Sequence 9, Appl  
40 1100.4 44.7 9709 2 US-08-188-583-5 Sequence 5, Appl  
41 1100.4 44.7 9709 3 US-08-388-353-1 Sequence 1, Appl  
42 1100.4 44.7 9709 3 US-08-488-551B-1 Sequence 1, Appl  
43 1100.4 44.7 9709 3 US-09-309-572-15 Sequence 15, Appl  
44 1100.4 44.7 9709 4 US-09-718-096-15 Sequence 15, Appl  
45 1100.4 44.7 12479 4 US-09-318-138-13 Sequence 13, Appl

## ALIGNMENTS

### RESULT 1

US-09-475-515-82  
; Sequence 82, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475.515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 82  
; LENGTH: 2306  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(-).protmod.NTopt.YM  
US-09-475-515-82

Query Match 83.1%; Score 2046; DB 4; Length 2306;  
Best Local Similarity 93.6%; Pred. No. 0;  
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;  
Qy 170 GCGCAAGGAGGGCCACCAGATGAAGGACTGCAGCGCGCCAGCCAACTTCTTCCGCG 229  
Db 1 GCGCGCGGAAAGGACACCAATGAAGATTGCATGAGACAGGCTAATTTCTTCCGCG 60  
Qy 230 AGGACCTGGCCCTCCCCCAAGGCGCCGAGTTCCTCCAGCGAGCAGAACCCGCGCA 289  
Db 61 AGGACCTGGCCCTCTTCGAGGCAAGCCCGCGAGTTTCAGCAGCGAGCAGACCCGCGCA 120  
Qy 290 ACAGCCCCACAGCCGCGAGCTGCAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGG 343  
Db 121 ACAGCCCCACCGCGCGAGCTGCAGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 180  
Qy 344 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 397  
Db 181 CCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

Qy 398 GCCCTGGTGAGCATCAAGTGGCGCGCAGATCAAGGAGGCCCTGCTGCAACACCGCG 457  
Db |||||  
241 GCCCTTGGTGAACATCAGGATCGCGCGCCAGCTCAAGGAGCGCTGCTGCAACCGCGCG 300  
Qy |||||  
458 CCAGCAGACACCGTCTGGAGAGATGAGCTGCCCGGCAAGTGGAAAGCCCAAGATGATCG 517  
Db |||||  
301 CCAGCAGACACCGTCTGGAGAGATGAACCTGCCCGGCAAGTGGAAAGCCCAAGATGATCG 360  
Qy |||||  
518 GCGGCATCGCGGGCTTCATCAAGGTGCGCCAGTAGTACGACAGATCTGATCGAGATCTGCG 577  
Db |||||  
361 GCGGGATCGGGGGCTTCATCAAGGTGCGGAGTAGTACGACAGATCTCCCGTGGAGATCTGCG 420  
Qy |||||  
578 GCAAGAGGCGCATCGGCACCGTGTGATCGGCCCGCCACCCCGTGAAACATCATCGGCCGCA 637  
Db |||||  
421 GCACAAGGCCATCGGCACCGTGTGTTGGGCCCGCCACCCCGTGAAACATCATCGGCCGCA 480  
Qy |||||  
638 ACATGTGACCCAGCTGGGCTGCAACCTGAACTTCCCATCAGCCCCATCGAGACCGTGC 697  
Db |||||  
481 ACCTGTGACCCAGATCGGCTGCACCTGAACCTTCCCATCAGCCCCATCGAGACCGTGC 540  
Qy |||||  
698 CCGTGAAGCTGAAGCCCGGCATGGAACGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 757  
Db |||||  
541 CCGTGAAGCTGAAGCCCGGCATGGAACGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 600  
Qy |||||  
758 AGAAGATCAAGGCCCTGACCGCATCTCGAGAGAGATGGAGAAGGGGCAAGATCACCA 817  
Db |||||  
601 AGAAGATCAAGGCCCTGTTGGAGATCTTGACCGGATGGAGAGGAGGCAAGATCAGCA 660  
Qy |||||  
818 AGATCGSCCCCGAAGACCCCTTACAACACCCCGTGTTCGCCATCAAGAAAGGACAGCA 877  
Db |||||  
661 AGATCGSCCCCGAAGACCCCTTACAACACCCCGTGTTCGCCATCAAGAAAGGACAGCA 720  
Qy |||||  
878 CCAAGTGGCGAAGCTGGTGGACTTCCGAGCTGAACAAGCGGACCCAGGACTTCTGGG 937  
Db |||||  
721 CCAAGTGGCGAAGCTGGTGGACTTCCGAGCTGAACAAGCGGACCCAGGACTTCTGGG 780  
Qy |||||  
938 AGGTGAGCTGGGCATCCCGACCCCGCGCTGAAGAAGAGAGAGGCTGACCGTGC 997  
Db |||||  
781 AGGTGAGCTGGGCATCCCGACCCCGCGCTGAAGAAGAGAGAGGCTGACCGTGC 840  
Qy |||||  
998 TGGAAGTGGCGAGCGCTACTTACGCTGCGCCCTGGACGAGGACTTCCGCAAGTACACCG 1057  
Db |||||  
841 TGGAAGTGGCGAGCGCTACTTACGCTGCGCCCTGGACGAGGACTTCCGCAAGTACACCG 900  
Qy |||||  
1058 CTTTACCATCCCGAGATCAACAAAGAGACCCCGCATCCGCTACAGTACAAAGTGC 1117  
Db |||||  
901 CTTTACCATCCCGAGATCAACAAAGAGACCCCGCATCCGCTACAGTACAAAGTGC 960  
Qy |||||  
1118 TGCCCCAGGGCTGGAAGGGCAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1177  
Db |||||  
961 TGCCCCAGGGCTGGAAGGGCAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCCTGG 1020  
Qy |||||  
1178 AGCCCTTCCGCGCCGCAACCCCGAGATCGTGTACTACAGGCCCCCTGTGACGTGGGCA 1237  
Db |||||  
1021 AGCCCTTCCGCAAGCAGAAACCCCGAGATCGTGTACTACAGGCCCCCTGTGACGTGGGCA 1080  
Qy |||||  
1238 GCGACTGGAGATCGGCCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCATCTGCTGC 1297  
Db |||||  
1081 GCGACTGGAGATCGGCCAGCACCGCGCAAGATCGAGAGCTGCGCGAGCATCTGCTGC 1140  
Qy |||||  
1298 GCTGGGGCTTCCACACCCCGCAAGAAAGCAACAGAGAGGAGCCCGCTTCTGTGGATGG 1357  
Db |||||  
1141 GCTGGGGCTTCCACACCCCGCAAGAAAGCAACAGAGAGGAGCCCGCTTCTGTGGATGG 1200  
Qy |||||  
1358 GCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGGCCCATCGAGCTGCCGAGAAAGGAGA 1417  
Db |||||  
1201 GCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGGCCCATCATGCTGCTGCCGAGAAAGCA 1260  
Qy |||||  
1418 GCTGGACCGTGAAGCATCCAGAGCTGGTGGGCAAGCTGAAGTGGGCCGAGCAGATCT 1477  
Db |||||  
1261 GCTGGACCGTGAAGCATCCAGAGCTGGTGGGCAAGCTGAAGTGGGCCGAGCAGATCT 1320  
Qy |||||  
1478 ACCCGGCATCAAGGTGCGGCAGCTGTGCAAGCTGTGCGCGGCCCAAGGCCCTGACCG 1537

Db |||||  
1321 ACCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCACCAAGGCCCTGACCG 1380  
Qy |||||  
1538 ACATGTGTCCTTGAACCGAGGAGCGAGCTGAGCTGGCCGAGAACCGCGAGATCTGTC 1597  
Db |||||  
1381 AGGTGATCCCTTGAACCGAGGAGCGAGCTGAGCTGGCCGAGAACCGCGAGATCTGTA 1440  
Qy |||||  
1598 GCGAGCCCGTGCAGCGGTGTACTAGACCCAGCAAGGACCTGTGCGCGAGATCCAGA 1657  
Db |||||  
1441 AGAGCCCGTGCACGAGGTGTACTAGACCCAGCAAGGACCTGTGTCGCGAGATCCAGA 1500  
Qy |||||  
1658 AGCAGGCGCACAGCAGTGGACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTGAAGA 1717  
Db |||||  
1501 AGCAGGCGCACAGGCGCAGTGGACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTGAAGA 1560  
Qy |||||  
1718 CCGCAGATGACGCAAGATGCGCACCGCCACACCAAGCAGTGAAGCAGTGAAGCAGG 1777  
Db |||||  
1561 CCGCAGATGACGCGCGCATGCGCGGCGCCACCAACGACGCTGAAGCAGTGAAGCAGG 1620  
Qy |||||  
1778 CCGTGCAGAGATGCGCATGCGAGAGATCGTGTGATCTGGGGCAAGACCCCAAGTTCGCGC 1837  
Db |||||  
1621 CCGTGCAGAGATGAGCACCGAGAGATCGTGTGATCTGGGGCAAGATCCCAAGTTCGAAGC 1680  
Qy |||||  
1838 TGCCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGCGAGGCCACCTGGA 1897  
Db |||||  
1681 TGCCCATCCAGAAAGGAGACCTGGGAGCGCTGGTGGATGGAGTACTGGCAGGCCACCTGGA 1740  
Qy |||||  
1898 TCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGTTACAGCTGAGGA 1957  
Db |||||  
1741 TCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGTTACAGCTGAGGA 1800  
Qy |||||  
1958 AGAGGCCCATCATCGCGCGCGAGACCTTCTACGTGGAACGCGCGCCCAACCGCGAGACCA 2017  
Db |||||  
1801 AGAGGCCCATCATCGCGCGCGAGACCTTCTACGTGGAACGCGCGCCCAACCGCGAGACCA 1860  
Qy |||||  
2018 AGATCGCGCAAGGCGCGCTACGTGACCGACCGCGGCGCGCAGAAAGATCGTGAGCTGACCG 2077  
Db |||||  
1861 AGCTGGCGAAGGCGCGCTACGTGACCGACCGCGGCGCGCAGAAAGTGTGTGAGCATCGCGC 1920  
Qy |||||  
2078 AGACCAACCAAGAGACCGAGCTGCGAGCGCATCCAGCTGGGCCCTGCGAGACAGCGGCA 2137  
Db |||||  
1921 ACACCAACCAAGAGACCGAGCTGCGAGCGCATCCACCTGGGCCCTGCGAGACAGCGGCC 1980  
Qy |||||  
2138 GCGAGGTGAACATCTGACCGACAGCAGTACCGCTGGGCATCATCCAGGCCCGAGGCCG 2197  
Db |||||  
1981 TGAAGTGAACATCTGTAACCAAGCAGCATGACCGCTGGGCATCATCCAGGCCCGAGGCCG 2040  
Qy |||||  
2198 ACAAGAGCGAGAGCGAGCTGGTGAACCAAGATCATTCGAGCAGCTGATCAAGAGGAGAGG 2257  
Db |||||  
2041 ACAAGAGCGAGAGCGAGCTGGTGAACCAAGATCATTCGAGCAGCTGATCAAGAGGAGAGG 2100  
Qy |||||  
2258 TGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGCGCAAGGACAGATCGACAGC 2317  
Db |||||  
2101 TGTACCTGGCTGGGTGCCCGCCCAAGGGCATCGCGCGCAAGGACAGTGGTGGACAGC 2160  
Qy |||||  
2318 TGTGAGCAAGGCGCATCCGCAAGTGTCTTCTGGAACGCGCATCGATGGCGGATCCGTA 2377  
Db |||||  
2161 TGTGAGCGCGCGCATCCGCAAGTGTCTTCTGGAACGCGCATCGATGGCGGATCCGTA 2220  
Qy |||||  
2378 TCTACCAAGTACATGGAACGACTGTGATCGTGGGCAAGCGGCCCTTAGGATCGATTTAAAGC 2437  
Db |||||  
2221 TCTACCAAGTACATGGAACGACTGTGATCGTGGGCAAGCGGCCCTTAGGATCGATTTAAAGC 2280  
Qy |||||  
2438 TTCCCGGGCTAGCACCGGTGAATTC 2463  
Db |||||  
2281 TTCCCGGGCTAGCACCGGTGAATTC 2306

## RESULT 2

US-09-475-515-84  
; Sequence 84, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:





```
Db 1741 CCTGGATCCCGAGTGGAGTTCGTGAACACCCCCCTGGTGAAGCTGTGGTACCAGC 1800
Qy 1952 TGGAGAAGGAGCCCATCATCGGGCCGAGACCTTCTAGTGGAGGGCGGCCCAACCGCG 2011
Db 1801 TGGAGAAGGAGCCCATCGTGGGGCCGAGACCTTCTACGTGGACGGCGGCCCAACCGCG 1860
Qy 2012 AGACCAAGATCGCAAGCGCGGTACGTGACCGACCGGGCCCGCAGAAAGATCGTAGCC 2071
Db 1861 AGACCAAGCTGGGCAAGCCCGCTACGTGACCGACCGGGCCCGCAGAAAGTGGTAGCA 1920
Qy 2072 TGACCGAGACACCAACCAAGAACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACA 2131
Db 1921 TCGCCGACACCAACCAAGAACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA 1980
Qy 2132 GCGGACGAGAGTGAACATCGTGACCGACAGCCAGTGCAGGCCATCCACCTGGCCCTGCAGGACC 2191
Db 1981 GCGGCTTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCAGGCC 2040
Qy 2192 AGCCGACAGAGCGAGCGAGCTGTGTGAACCAAGGCATCGAGCAGCTGATCAAGAAGG 2251
Db 2041 AGCCCGACAAGAGCGAGCGAGCTGTGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG 2100
Qy 2252 AGAAGTGTACCTGAGCTGGGTGCCGCCACCAAGGCATCGCGGCCCAACGAGCAGATCG 2311
Db 2101 AGAAGTGTACCTGGCTGGTGGCCGCCACCAAGGCATCGAGCAGCTGATCAAGAAGG 2160
Qy 2312 ACAAGCTGTGAGCAAGGGGATCGGCAAGTGCTGTTCCTGGACGGGATCGATGGGGCA 2371
Db 2161 ACAAGCTGTGAGCGCGCGGATCGCAAGGTGTGTTCCTGAACGGGATCGATGGGGCA 2220
Qy 2372 TCGTGATCTACAGTACATGAGCAGCTGTACGTGGGCGGCGGCGGCTAGGATCGATT 2431
Db 2221 TCGTGATCTACAGTACATGAGCAGCTGTACGTGGGCGGCGGCGGCGGCGGCGGCGG 2280
Qy 2432 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2463
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312
```

```
RESULT 3
US-09-475-515-83
; Sequence 83, Application US/09475515A
; Patent No. 6602705
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: ZUR MEGEDE, Jan
; APPLICANT: SRIVASTAVA, Indresh
; APPLICANT: LIAN, Ying
; APPLICANT: LIAN, Ying
; APPLICANT: HARTOG, Karin
; APPLICANT: LIU, Hong
; APPLICANT: GREER, Catherine
; APPLICANT: SELBY, Mark
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES
; FILE REFERENCE: 1621.002
; CURRENT APPLICATION NUMBER: US/09/475.515A
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 83
; LENGTH: 2300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: FS(-).protmod.rTopt.YMMH
US-09-475-515-83
```

```
Query Match 82.0%; Score 2019.2; DB 4; Length 2300;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
```

```
Qy 170 GCGCAAGGAGGCGCCACAGATGAAGACTGACCGAGCGCCAGGCGCAACTTCTTCGCG 229
Db 1 GCGGCGCGNAGAGACACCAANTGAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60
Qy 230 AGGACCTGGCTTTCCCGCAGGGCAAGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCCA 289
Db 61 AGGACCTGGCTTTCTGCGAGGGCAAGCGCGCGAGTTTCAGCAGCGAGCAGACCGCGCCA 120
Qy 290 ACAGCCCCCAGCGCGGAGCTCGAGTGGCGGGG-----ACNACCCCGCAGCGAGG 343
Db 121 ACAGCCCCCAGCGCGGAGCTCGAGTGTGGGCGGCGAGAACACAGCCTTGAGCGAGG 180
Qy 344 CCGGCGCGAGCGCCAGGCGACCTG-----AATTTCCCGCAGATCACCTGTGGCAGC 397
Db 181 CCGGCGCGAGCGCGAGGCGACCGTGAGCTTCAACTTTCCCGCAGATCACCTGTGGCAGC 240
Qy 398 GCGCCCTGTGTAGCATCAAGGTGGCGCGCAGATCAAGGAGGCGCTGTCTGGACACCGCG 457
Db 241 GCGCCCTGTGTAGCATCAGGATCGGCGCGCAGCTCAAGGAGGCGCTGTCTGACACCGCG 300
Qy 458 CCGACGACACCGTGTGTGAGGAGATGAGCTGCGCGGCAAGTGGAGAGCCCAAGATGATCG 517
Db 301 CCGACGACACCGTGTGTGAGGAGATGAACCTGCGCGGCAAGTGGAGAGCCCAAGATGATCG 360
Qy 518 GCGGATCGGCGCGCTTTCATCAAGGTGCGCGCAGTACGACAGATCCTGTGATCGAGATCTGCG 577
Db 361 GCGGATCGGCGCGCTTTCATCAAGGTGCGCGCAGTACGACAGATCCTGTGAGATCTGCG 420
Qy 578 GCAAGAGGCGCATCGGCGCGCTGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637
Db 421 GCGCAAGGCGCATCGGCGCGCTGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCA 480
Qy 638 ACATGCTGACCGAGCTGGGCTGCACCTGAACTTTCCCGCATCAGCGCGCGCGCGCGCG 697
Db 481 ACCTGCTGACCGAGATCGGCTGCACCTGAACTTTCCCGCATCAGCGCGCGCGCGCGCG 540
Qy 698 CCGTGAAGCTGAAGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
Db 541 CCGTGAAGCTGAAGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Qy 758 AGAAGATCAAGGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 817
Db 601 AGAAGATCAAGGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Qy 818 AGATCGGCGCGGAGACCCCTTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
Db 661 AGATCGGCGCGGAGACCCCTTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Qy 878 CCAAGTGGCGCAAGCTGGTGAAGTTCGCGAGCTGAAACAGCGCACCGCAGGACTTCTGGG 937
Db 721 CCAAGTGGCGCAAGCTGGTGAAGTTCGCGAGCTGAAACAGCGCACCGCAGGACTTCTGG 780
Qy 938 AGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 997
Db 781 AGGTGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Qy 998 TGAAGTGGCGGAGCGCTACTTTCAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057
Db 841 TGAAGTGGCGGAGCGCTACTTTCAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Qy 1058 CTTTCAACATCCCGCAGCATCAACAGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1117
Db 901 CTTTCAACATCCCGCAGCATCAACAGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Qy 1118 TGCCCCAGGCGCTGGAAGGCGAGCCCGCAGCATCTTTCAGAGCAGCATGACCAAGATCTGG 1177
Db 961 TGCCCCAGGCGCTGGAAGGCGAGCCCGCGCATCTTTCAGAGCAGCATGACCAAGATCTGG 1020
Qy 1178 AGCCCTTCGCGCGCGCGCGAGATCGTGATCTTACAGGCGCGCGCGCGCGCGCGCGCGCG 1237
Db 1021 AGCCCTTCGCGCGCGCGCGAGAACCCCGGACATGTGTACTACAGGCGCGCGCGCGCGCG 1080
```



QY 441 CTTGCTGGACACCGGCGCCGACACACCGTCTGGAGGAGATGAGCCTGCGCCGCAAGTG 500  
DB 1541 GCTGCTGCGACACCGGCGCCGACACACCGTCTGGAGGAGATGAACTGCGCCGCAAGTG 1600  
QY 501 GAAGCCCAAGATGATCGCGGGCATCGCGGCTTTCATCAAGGTGCGCCAGTACGACAGAT 560  
DB 1601 GAAGCCCAAGATGATCGCGGGATCGGGGCTTTCATCAAGGTGCGCAGTACGACAGAT 1660  
QY 561 CTTGATGAGATCTGGCGCAAGAGGCAATCGGCAACCGTCTGATCGGCGCCCAACCCCGT 620  
DB 1661 CCCCGTGGAGATCTGGCGCCACAAAGGCAATCGGCAACCGTCTGCTGGTGGCGCCCAACCCCGT 1720  
QY 621 GAACATCATCGGCGCAACATGCTGACCCAGATCGGCTGACCCCTGAACTTCCCAATCAG 680  
DB 1721 GAACATCATCGGCGCAACATGCTGACCCAGATCGGCTGACCCCTGAACTTCCCAATCAG 1780  
QY 681 CCCCATCGAGACCGTCCCGTGAAGCTGAAAGCCCGGATGAGCGGCGCCCAAGGTGAAGCA 740  
DB 1781 CCCCATCGAGACCGTCCCGTGAAGCTGAAAGCCCGGATGAGCGGCGCCCAAGGTGAAGCA 1840  
QY 741 GTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGGAGGAGATGGAGAA 800  
DB 1841 GTGGCCCTGACCGAGAGAGATCAAGGCCCTGAGGCTGAGATCTGCAACGAGATGGAGAA 1900  
QY 801 GGAGGCAAGATCAACCAAGATCGGCGCGGAGAACCCCTCAACACACCCCGTGTTCGCCAT 860  
DB 1901 GGAGGCAAGATCAGCAGAGATCGGCGCGGAGAACCCCTCAACACACCCCGTGTTCGCCAT 1960  
QY 861 CAAGAAGAAGGACAGCACCAAGTGGGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 920  
DB 1961 CAAGAAGAAGGACAGCACCAAGTGGGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCG 2020  
QY 921 CACCCAGGACTTCTGGAGAGTGAGCTGGGCAATCCGCCACCCCGCGGCTTGAAGAAGAA 980  
DB 2021 CACCCAGGACTTCTGGAGAGTGAGCTGGGCAATCCGCCACCCCGCGGCTTGAAGAAGAA 2080  
QY 981 GAAGAGGTCACCGTGTGACCTGGGCGAGCGCTACTTTCAGAGTGCCCTGACGAGGA 1040  
DB 2081 GAAGAGGTCACCGTGTGACCTGGGCGAGCGCTACTTTCAGAGTGCCCTGACGAGGA 2140  
QY 1041 CTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAGAGACCCCGGCAATCG 1100  
DB 2141 CTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAGAGACCCCGGCAATCG 2200  
QY 1101 CTACCAAGTACACCGTGTGCGGCTGGAGGCGAGCCCGAGCATCTTTCAGAGCAG 1160  
DB 2201 CTACCAAGTACACCGTGTGCGGCTGGAGGCGAGCCCGAGCATCTTTCAGAGCAG 2260  
QY 1161 CATGACCAAGATCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCTACCA -- 1217  
DB 2261 CATGACCAAGATCTGGAGCCCTTCCGCGCGAGTACGAGCTGCAACCCCGAGATCGTGATCTACCA 2320  
QY 1218 ---GGGCCCCCTGTACGTGGGCGAGCGACTGGAGATCGGCGAGCAGCCGCGCAAGATCGA 1274  
DB 2321 CATGGAGGACCTGTACGTGGGCGAGCGACTGGAGATCGGCGAGCAGCCGCAACCAAGATCGA 2380  
QY 1275 GGAGCTGCGCAAGCACTGTGCTGGGCTTCAACACCCCGCAAGAGACCAAGAA 1334  
DB 2381 GGAGCTGCGCAAGCACTGTGCTGGGCTTCAACACCCCGCAAGAGACCAAGAA 2440  
QY 1335 GGAGCCCGCTTCTGTGGATGGCTACGAGCTGCAACCCCGCAAGTGGAGCCGTGCAAGCC 1394  
DB 2441 GGAGCCCGCTTCTGTGGATGGCTACGAGCTGCAACCCCGCAAGTGGAGCCGTGCAAGCC 2500  
QY 1395 CATCGAGCTCGCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAA 1454  
DB 2501 CATCGATCTCGCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAA 2560  
QY 1455 GCTGAATCGGCGCAGCAGATCTACCCCGGATCAAGGTGGCGGCTGTGCAAGCTGCT 1514  
DB 2561 GCTGAATCGGCGCAGCAGATCTACCCCGGATCTACCCCGGATCAAGGTGAAGCAGCTGTGCAAGCTGCT 2620

QY 1515 GCSCGCGCCCAAGGCGCCCTGACCGACATCGTGCCCTGACCGAGGAGGCGGAGCTGGAGCT 1574  
DB 2621 GCSCGCGCCCAAGGCGCCCTGACCGAGCTGATCCCTCTGACCGAGGAGGCGGAGCTGGAGCT 2680  
QY 1575 GGCCGAGAACCGCGAGATCTTGGCGAGCCCTGTCAGCGGCTGTACTACGACCCGAGCAA 1634  
DB 2681 GGCCGAGAACCGCGAGATCTTGAAGGAGCCCTGTGCAAGGCTGTACTACGACCCGAGCAA 2740  
QY 1635 GGACCTGTGGCGCGAGATCCAGAGCAGGCGCCAGCAGCTGAGCTTACAGATCTTACCA 1694  
DB 2741 GGACCTGTGGCGCGAGATCCAGAGCAGGCGCCAGGCGCCAGTGGACCTTACAGATCTTACCA 2800  
QY 1695 GGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCAAGATCGCACCCGCGAGCAA 1754  
DB 2801 GGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCAAGATCGCGCGCGCGAGCAA 2860  
QY 1755 CGACGTGAAGCAGCTGACCGAGCGCTGCAAGAGATCGCCATCGAGAGCATCTGTGATCTG 1814  
DB 2861 CGACGTGAAGCAGCTGACCGAGCGCTGCAAGAGATCGCCATCGAGAGCATCTGTGATCTG 2920  
QY 1815 GGCGAGAGCCCAAGTTCGCTGCGCTGCCATCCAGAGAGGAGACCTGGAGACCTTGGTGAC 1874  
DB 2921 GGCGAGAGATCCCAAGTTCGCTGCGCTGCCATCCAGAGAGGAGACCTGGAGAGCTTGGTGAT 2980  
QY 1875 CGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAACAACCCCGCGCTGGT 1934  
DB 2981 GGAGTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTAACAACCCCGCGCTGGT 3040  
QY 1935 GAAGCTGTGTGTACACGCTGGAGAGAGCCCATCATCTGGGCGCGAGACCTTCTACGTGGA 1994  
DB 3041 GAAGCTGTGTGTACACGCTGGAGAGAGCCCATCATCTGGGCGCGAGACCTTCTACGTGGA 3100  
QY 1995 CGSCGCGCCCAACCGGAGACCAAGATCGCAAGGCGGCTAGCTGACCGACCGGGGCGG 2054  
DB 3101 CGSCGCGCCCAACCGGAGACCAAGCTGGCAAGGCGGCTAGCTGACCGACCGGGGCGG 3160  
QY 2055 GCAGAAGATCGTGAGCTGACCGAGACCAACCAAGAGAGCCGAGCTGAGGCGCATCCA 2114  
DB 3161 CCAGAGGTGGTGAGCATCGCCGACACCAACCAAGAGAGCCGAGCTGAGGCGCATCCA 3220  
QY 2115 GCTGGCCCTGAGGACAGCGGCGAGCGTGAACATCTGTGACCGAGCAGCGAGTACGCGCT 2174  
DB 3221 CTTGGCCCTGAGGACAGCGGCGTGGAGGTGAACATCTGTGACCGAGCAGCGAGTACGCGCT 3280  
QY 2175 GGSCATCATCCAGGCGCGAGCCGAGCAAGAGAGCGAGCTGTGAACAGATCATCGA 2234  
DB 3281 GGSCATCATCCAGGCGCGAGCCGAGCAAGAGAGCGAGCTGTGAACAGATCATCGA 3340  
QY 2235 GCAGCTGATCAAGAAGAGAGGTGTACTGTGAGCTGGTGCCCGCCCAAGGGGCAATCGG 2294  
DB 3341 GCAGCTGATCAAGAAGAGAGGTGTACTGTGAGCTGGTGCCCGCCCAAGGGGCAATCGG 3400  
QY 2295 CGGCAACGAGCAGATCGACAGCTGGTGGAGAGGGGATCCGCAAGGTGCTGTCTCTGGA 2354  
DB 3401 CGGCAACGAGCAGGTGGAGCAAGCTGGTGGAGCGCGGATCCGCAAGGTGCTGTCTCTGGA 3460  
QY 2355 CGSCATCATGGCGGATCGTGATCTACCACTA 2387  
DB 3461 CGSCATCATGGCGGATCGTGATCTACCACTA 3493

## RESULT 5

US-09-475-515-80  
; Sequence 80, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGHEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine

; APPLICANT: SELBY, Mark  
 ; APPLICANT: WALKER, Christopher  
 ; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
 ; FILE REFERENCE: 1621.002  
 ; CURRENT APPLICATION NUMBER: US/09/475,515A  
 ; CURRENT FILING DATE: 1999-12-30  
 ; NUMBER OF SEQ ID NOS: 90  
 ; SOFTWARE: Patent in Ver. 2.0  
 ; SEQ ID NO 80  
 ; LENGTH: 2305  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence:  
 ; OTHER INFORMATION: FS(+).proinact.RTopt.YM  
 US-09-475-515-80

Query Match 76.3%; Score 1878.8; DB 4; Length 2305;  
 Best Local Similarity 89.4%; Pred. No. 7.2e-312;  
 Matches 2061; Conservative 0; Mismatches 232; Indels 13; Gaps 3;

Qy	170	GCAGCAAGGAGGCGCCACAGATGAAGGACTGCACCGAGCGCGCAGCCAACTTCTTCGCGC	229
Db	1	CGGCGCGGAAGGACACCAATGAAGATTGCATGAGACAGGCTAAATTTT-AGGG	59
Qy	230	AGGACCTGGCTTCCCCCAGGCGCAAGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCA	289
Db	60	AAGATCTGGCTTCTTCAAGGGAAGCCACGGGAATTTTCTTCAGAGCAGACAGAGCA	119
Qy	290	ACAGCCCAACAGCGCGGAGCTGAGGTGCGCGGG-----ACAAACCCCGCAGCGAGG	343
Db	120	ACAGCCCAACAGAGAGAGCTTCAGGTTTGGGAGGAGAAACAACTCCCTCTCAGAA	179
Qy	344	CCGCGCGGAGCGCCAGGCA-----CCCTGAATCTCCCGCAGATCACTCTGGCGAGC	397
Db	180	CAGAGCGGATAGAACAGGAATCTGTATCTTTAACTTCTCCCTCAGATCACTCTTTGGCAAC	239
Qy	398	GCCCGCTGGTGAGATCAAGGTGGCGCGCCAGATCAAGGAGCGCTGTGGACACCGCGC	457
Db	240	GACCCCTCGTCACAATAAGGATCGGGGGCAACTCAAGGAAGCGTGTCTGATACAGGAG	299
Qy	458	CCGACGACACCGTCTGGAGAGATGAGCCTCGCGCGCAAGTGAAGCCCAAGATGATCG	517
Db	300	CAGATGATACAGTATTAGAAAGAAATGAAATTTGCCAGGAAATGGAACCAAAATGATAG	359
Qy	518	GCGCATCGCGGCTTCATCAAGGTGCGCAGTACGACCCAGATCCTGATCGAGATCTGCG	577
Db	360	GGGGATCGGGGGCTTCATCAAGGTGAGGCGAGTACGACCCAGATACCTGTGAAATCTGTG	419
Qy	578	GCAAGAAGGCGCATCGGCACCGTGTGATCGCGCCCGCCACCCCGTGAACATCATCGCGCGCA	637
Db	420	GACATAAGCTATAGGTACAGTATTAGTAGAACCTACACCTGTCAACATATTGGAAGAA	479
Qy	638	ACATGCTGACCCAGTGGGTGCAACCTGAATTTCCCATCAGCCCGCATCGAGACCGTGC	697
Db	480	ATCTGTTGACCCAGATCGGCTGCACCTTGAATTTCCCATCAGCCCTATTGAGACGGTGC	539
Qy	698	CCGTGAAGCTGAAGCCCGCATGAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGG	757
Db	540	CCGTGAAGTGAAGCCCGGGATGAGCGCCCAAGGTCAAGCAATGGCCATTGACCGAGG	599
Qy	758	AGAGATCAAGCCCTGACCCCATCTCGAGGAGATGGAGAAGGAGGGCAAGATCACCA	817
Db	600	AGAAGATCAAGCCCTGTTGAGATCTGACCGGAGATGGAGAGGGCAAGATCAGCA	659
Qy	818	AGATCGGCCCGAGAACCCCTTACCAACCCCGGTTTGGCCATCAAGAAGAGACAGCA	877
Db	660	AGATCGGCCCGAGAACCCCTTACCAACCCCGGTTTGGCCATCAAGAAGAGACAGCA	719
Qy	878	CCAAGTGGCGCAAGCTGGTGAATCTCGCGAGCTGAACCAAGCGCACCCAGACTTCTGGG	937
Db	720	CCAAGTGGCGCAAGCTGGTGAATCTCGCGAGCTGAACCAAGCGCACCCAGACTTCTGGG	779

Qy	938	AGGTGAGCTGGGCAATCCCCCAACCCCGCGCCCTGAAGAAAGAAAGAGCGTGAACCGTGC	997
Db	780	AGGTGAGCTGGGCAATCCCCCAACCCCGCGCCCTGAAGAAAGAAAGAGCGTGAACCGTGC	839
Qy	998	TGGACGTGGGCGAGCGCTTCTTCAAGCTGCGCCCTGGACGAGACTTCCGCAAGTACACCG	1057
Db	840	TGGACGTGGGCGAGCGCTTCTTCAAGCTGCGCCCTGGACGAGACTTCCGCAAGTACACCG	899
Qy	1058	CTTTCACCATCCCGCAGATCAACCAACGAGAGACCCCGCGCATCCGCTACCAAGTACACCGTGC	1117
Db	900	CTTTCACCATCCCGCAGATCAACCAACGAGAGACCCCGCGCATCCGCTACCAAGTACACCGTGC	959
Qy	1118	TGCCCCAGAGGCTGGAAGGCGAGCCCGCAGACTTCTTCAGAGCAGCATGACCAAGATCTCTGG	1177
Db	960	TGCCCCAGAGGCTGGAAGGCGAGCCCGCAGACTTCTTCAGAGCAGCATGACCAAGATCTCTGG	1019
Qy	1178	AGCCCTTCCGCGCGCAACCCCGCAGATCGTGATCTACCAAGGAGGAGGAGGAGGAGCA	1237
Db	1020	AGCCCTTCCGCGCGCAACCCCGCAGATCGTGATCTACCAAGGAGGAGGAGGAGGAGCA	1079
Qy	1238	GCAGCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGTCTGC	1297
Db	1080	GCAGCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGTCTGC	1139
Qy	1298	GCTGGGCTTCCACCAACCCCGCAAGAGCAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG	1357
Db	1140	GCTGGGCTTCCACCAACCCCGCAAGAGCAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG	1199
Qy	1358	GCTACGAGCTGCAACCGCGCAAGTGGAGCGCTGAGCCCATCATGCTGCTGCGCGAGGAGGAG	1417
Db	1200	GCTACGAGCTGCAACCGCGCAAGTGGAGCGCTGAGCCCATCATGCTGCTGCGCGAGGAGGAG	1259
Qy	1418	GCTGGAACCGTGAACCAACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCGAGTCT	1477
Db	1260	GCTGGAACCGTGAACCAACATCCAGAGCTGGTGGGCAAGCTGAACTGGGCGCAGCGAGTCT	1319
Qy	1478	ACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCG	1537
Db	1320	ACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCG	1379
Qy	1538	ACATGTCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGC	1597
Db	1380	AGGTGATCCCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGA	1439
Qy	1598	GCAGACCGTGCAGCGGCTGTACTACGACCCCAAGGAGCTGCTGGTGGCGCGAGATCCAGA	1657
Db	1440	AGGAGCCCGTGCAGCGGCTGTACTACGACCCCAAGGAGCTGCTGGTGGCGCGAGATCCAGA	1499
Qy	1658	AGCAGGCGCAACGACCTTACAGATCTTACAGGAGCGCTTCAAGAACCTGGAAGA	1717
Db	1500	AGCAGGCGCAACGACCTTACAGATCTTACAGGAGCGCTTCAAGAACCTGGAAGA	1559
Qy	1718	CCGCAAGTACCGCAAGATGGGCAACCGCGCACCAAGACGAGCTGAAGCAGTGAACCGAGG	1777
Db	1560	CCGCAAGTACCGCAAGATGGGCAACCGCGCACCAAGACGAGCTGAAGCAGTGAACCGAGG	1619
Qy	1778	CCGTGCAAGAGATCGCCATGAGAGAGCTGCTGATCTGGGGCAAGACCCCAAGTTCCGCC	1837
Db	1620	CCGTGCAAGAGATCGCCATGAGAGAGCTGCTGATCTGGGGCAAGATCCCCAAGTTCAAGC	1679
Qy	1838	TGCCCATCCAGAGGAGACCTTGGGAGACCTGGTGGACCGACTACTGGCAGGCGCACCTGGA	1897
Db	1680	TGCCCATCCAGAGGAGACCTTGGGAGGCTTGGTGGATGAGTACTTGGCAGGCGCACCTGGA	1739
Qy	1898	TCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACAGCTGGAGA	1957
Db	1740	TCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACAGCTGGAGA	1799
Qy	1958	AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGCGCGCCCAACCGCGAGACCA	2017
Db	1800	AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGCGCGCGCCCAACCGCGAGACCA	1859

QY 2018 AGATCGCAAGCGCGCTACGTGACCGACCGGGCGGCAGAAAGATCGTGAGCTGACCG 2077  
|||  
Db 1860 AGCTGGCAAGCGCGCTAGCTGACCGACCGGGCGGCAGAAAGGTGGTGAGCATCGCCG 1919  
|||  
QY 2078 AGACCAACCAAGAGACCGAGCTGCGAGGCCATTCAGAGCTGGCCCTGCGAGGACAGCGGCA 2137  
|||  
Db 1920 ACACCAACCAAGAGACCGAGCTGCGAGGCCATTCACCTGGCCCTGCGAGGACAGCGGCC 1979  
|||  
QY 2138 GCGAGGTGAACATCGTGACCGACGCGAGTACGCCCTGGCGCATCATCCAGGCCCGACGCCG 2197  
|||  
Db 1980 TGGAGGTGAACATCGTGACCGACGCGAGTACGCCCTGGCGCATCATCCAGGCCCGACGCCG 2039  
|||  
QY 2198 ACAAGAGCGAGCGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2257  
|||  
Db 2040 ACAAGAGCGAGCGAGCTGGTGAACCGAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG 2099  
|||  
QY 2258 TGTACTGAGCTGGGTGCCGCCACCAAGGGCATTCGGCGCAACGAGCAGATCGACAAGC 2317  
|||  
Db 2100 TGTACTGGCTGGGTGCCGCCACCAAGGGCATTCGGCGCAACGAGCAGTGGACAAGC 2159  
|||  
QY 2318 TGGTGAGCAAGGCATCCGCAAGGTGCTGTTCTGGACGGCATCGATGGCGGCATCGTGA 2377  
|||  
Db 2160 TGGTGAGCGCGGCATCCGCAAGGTGCTGTTCTGGAACGGCATCGATGGCGGCATCGTGA 2219  
|||  
QY 2378 TCTACAGTACATGGAGCAGCTGTACGTGGCGAGCGCGGCCCTAGGATCGATTTAAAGC 2437  
|||  
Db 2220 TCTACAGTACATGGAGCAGCTGTACGTGGCGAGCGCGGCCCTAGGATCGATTTAAAGC 2279  
|||  
QY 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463  
|||  
Db 2280 TTCCCGGGGCTAGCACCGGTGAATTC 2305  
|||

## RESULT 6

US-09-475-515-81  
; Sequence 81, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEDEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; TITLE OF INVENTION: OF VIRUS-LIKE PARTICLES  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 81  
; LENGTH: 2299  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(+).proinaet.RTopt.YMMW  
US-09-475-515-81

Query Match 75.2%; Score 1852; DB 4; Length 2299;  
Best Local Similarity 89.0%; Pred. No. 2.7e-307;  
Matches 2052; Conservative 0; Mismatches 235; Indels 19; Gaps 4;

QY 170 GCGCAAGGAGGCCACCAAGTGAAGGACTGCAACGAGCGCCAGGCCAACTTCTTCGCG 229  
|||  
Db 1 GCGGCGCGAAGGACACCAATGAAGATTGCACTGAGACAGGCTTAATTTTT-AGGG 59  
|||  
QY 230 AGGACCTGGCTTTCCCGCAGGCGAAGCCCGCGAGTTTCCCGAGCGAGCAAGCGGCCA 289  
|||

Db 60 AAGATCTGGCTTCTTCAAGGGAAGCGCCAGGGAAATTTTCTTCAGAGCAGACCGAGGCCA 119  
QY 290 ACAGCCCCACCGAGCGCGAGCTGCAGGTGGCGGG-ACAAACCCCGCAGCGAGG 343  
|||  
Db 120 ACAGCCCCACCGAGAGAGCTTCAGGTTTGGGAGGAGAAACAATCTCCTCTCAGAAG 179  
|||  
QY 344 CCGGCGCCGAGCGCCAGGGCA-----CCCTGAACCTTCCCGCAGATCACCTGTGGCAGC 397  
|||  
Db 180 CAGGAGCCGATAGACAAGGAACGTATCTTTAACTTCCCTCAGATCACTCTTTGGCAAC 239  
|||  
QY 398 GCCCCTGTGTAGCATCAAGGTGGCGGCCAGATCAAGGAGGCCCTGTCTGGACACCGGCG 457  
|||  
Db 240 GACCCCTCGTCAATAAAGGATCGGGGGCACTCAAGGAAGCGCTGCTCGATACAGGAG 299  
|||  
QY 458 CCGACGACACCGTGTGGAGGAGATGAGCTGCCCGCAAGTGAAGGCCCAAGATGATCG 517  
|||  
Db 300 CAGATGATACAGTATTAGAAAGAAATGAAATTGTCAGGAAATGAAACCAAAATGATAG 359  
|||  
QY 518 GCGGCATCGCGCGCTTTCATCAAGGTGCGCCAGTACGACACAGATCTCTGATCGAGATCTGCG 577  
|||  
Db 360 GGGGATCGGGGCTTTCATCAAGGTGAGCGAGTACGACACAGATACCTGTAGAAATCTGTG 419  
|||  
QY 578 GCAAGAGGCGCATCGGCGACCGTCTGATCGGCCCGACCCCGGTGAACATCATCGGCCGCA 637  
|||  
Db 420 GACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAAATTGGAGAA 479  
|||  
QY 638 ACATGTCGACCGAGCTGGGCTGCACCTGAACTTCCCGCATCAGCCCATCGAGACCGTGC 697  
|||  
Db 480 ATCTGTGACCGCAGATCGGCTGCACCTTGAACCTTCCCGCATCAGCCCTTATTTGAGACGGTGC 539  
|||  
QY 698 CCGTGAAGCTGAAGCGCGCATGGAAGCGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGG 757  
|||  
Db 540 CCGTGAAGTTGAAGCGCGGATGGAAGCGCGCCCAAGGTCAAGCAATGGCCATTGACCGAGG 599  
|||  
QY 758 AGAAGATCAAGGCTTGAACCGCATCTGCGAGGAGATGGAAGAGGAGGGCAAGATCAACA 817  
|||  
Db 600 AGAAGATCAAGGCTTGGTGGAGTCTGCACCGAGATGGAAGAGGAGGGCAAGATCAAGCA 659  
|||  
QY 818 AGATCGGCGCGAGAACCCCTTCAACACCCCGCTTCCGCCATCAAGAAGAAGACAGCA 877  
|||  
Db 660 AGATCGGCGCGAGAACCCCTTCAACACCCCGCTTCCGCCATCAAGAAGAAGACAGCA 719  
|||  
QY 878 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG 937  
|||  
Db 720 CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGG 779  
|||  
QY 938 AGTGAGCTGGGCGATCCCGACCCCGCGCTTGAAGAAGAGAGAGGCGTCAACCGTGC 997  
|||  
Db 780 AGTGAGCTGGGCGATCCCGACCCCGCGCTTGAAGAAGAGAGAGGCGTCAACCGTGC 839  
|||  
QY 998 TGGACGTGGCGAGCGCTACTTTCAGCGTGGCCCTTGGACGAGGACTTCCGCAAGTACACCG 1057  
|||  
Db 840 TGGACGTGGCGAGCGCTACTTTCAGCGTGGCCCTTGGACGAGGACTTCCGCAAGTACACCG 899  
|||  
QY 1058 CTTTCACCATCCCGCAGCATCAACAGAGACCCCGCGCATCCGCTACAGTACAACTGC 1117  
|||  
Db 900 CTTTCACCATCCCGCAGCATCAACAGAGACCCCGCGCATCCGCTACAGTACAACTGC 959  
|||  
QY 1118 TGCCCGAGGCTGGAGGGGAGCGCCCGAGCATCTTTCAGAGCAGCATGACCAAGATCTGG 1177  
|||  
Db 960 TGCCCGAGGCTGGAGGGGAGCGCCCGCGCATCTTTCAGAGCAGCATGACCAAGATCTCG 1019  
|||  
QY 1178 AGCCCTTCGCGCGCCCAACCCCGAGATCGTGATCTACAGGCGCCCGCTGAGTGGGCA 1237  
|||  
Db 1020 AGCCCTTCGCGCAAGAGAGACCCCGAGCATCTGATCTACAGGCGCCCGCTGATCGTGGGCA 1079  
|||  
QY 1238 GCGACCTTGAAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCAAGCACTGCTGC 1297  
|||  
Db 1080 GCGACCTTGAAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCGACCACTGCTGC 1139  
|||  
QY 1298 GCTGGGGCTTCAACACCCCGCAAGAGCACCAAGAGGAGCGCCCGCTTCTGTGGATGG 1357  
|||  
Db 1140 GCTGGGGCTTCAACACCCCGCAAGAGCACCAAGAGGAGCGCCCGCTTCTGTCCCAT-- 1197  
|||

```
Qy 1358 GCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGTGCCTCCCGAGAGGAGA 1417
Db 1198 ----CGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCATGCTGCTCCCGAGAGGACA 1253
Qy 1418 GCTGAGCCGTGAACGACATCCAGAACCTGGTGGGCAAGCTGAACTGGGSCCAGCCAGATCT 1477
Db 1254 GCTGAGCCGTGAACGACATCCAGAACCTGGTGGGCAAGCTGAACTGGGSCCAGCCAGATCT 1313
Qy 1478 ACCCGGCGATCAAGGTGCGCAGCTGTGCAAGCTGTGCGGGCGCCAAAGGCCCTGACCG 1537
Db 1314 ACGCCGCGATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGGGCGCCAAAGGCCCTGACCG 1373
Qy 1538 ACATGTCCTCTGACCGAGAGGCGGAGCTGGAAGCTGGGCCGAGAAACCGGAGATCTCTGC 1597
Db 1374 AGTGATCCCCCTGACCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGA 1433
Qy 1598 GCAGCCCGTGCACGGCGTGTACTACGACCCAGACCCAGGACCTGTGTGCGCGAGATCCAGA 1657
Db 1434 AGGAGCCCGTGCAAGAGTGTACTACGACCCAGACCCAGGACCTGTGTGCGCGAGATCCAGA 1493
Qy 1658 AGCAGGCGCAGCAGCAGTGGACCTTACCAGATCTTACCAGGAGCCCTTCAAGAACCTGAAGA 1717
Db 1494 AGCAGGCGCAGGCGCAGTGGACCTTACCAGATCTTACCAGGAGCCCTTCAAGAACCTGAAGA 1553
Qy 1718 CCGCAAGTACCGCAAGATGGCGACCGGCCACACCAACGACGTGAAGAGCTGACCGAGG 1777
Db 1554 CCGCAAGTACCGCCGATGCGCGCGGCCACACCAACGACGTGAAGAGCTGACCGAGG 1613
Qy 1778 CCGTGCAAGAGATCGCCATGAGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCC 1837
Db 1614 CCGTGCAAGAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC 1673
Qy 1838 TGCCCATTCGAAGAGAGACCTGGGAGACCTGGTGGACCGAGTACTGGCAGGCCACCTGGA 1897
Db 1674 TGCCCATTCGAAGAGAGACCTGGGAGGCTTGGTGGATGAGTACTGGCAGGCCACCTGGA 1733
Qy 1898 TCCCGAGTGGAGTTCGTGAACACCCGCCCTTGGTGAAGCTGTGTGTAACAGCTGGAGA 1957
Db 1734 TCCCGAGTGGAGTTCGTGAACACCCGCCCTTGGTGAAGCTGTGTGTAACAGCTGGAGA 1793
Qy 1958 AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGCGCCGCAACCGCGAGACCA 2017
Db 1794 AGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGCGCCGCAACCGCGAGACCA 1853
Qy 2018 AGATCGGCAAGCGCGGCTACGTGACCGACCGGGGCGCGCAGAGATCGTGAGCCTGACCG 2077
Db 1854 AGCTGGGCAAGCGCGGCTACGTGACCGACCGGGGCGCGCAGAGAGGTGTTGAGCATCGCG 1913
Qy 2078 AGACCAACCAAGAGACCGAGTGCAGGCCATCCAGCTGGCCCTGCGAGGACAGCGGCA 2137
Db 1914 ACACCAACCAAGAGACCGAGTGCAGGCCATCCAGCTGGCCCTGCGAGGACAGCGGCG 1973
Qy 2138 GCGAGGTGACATCGTGACCGAGACCGAGTACGCGCTGGGCGATCATCCAGGCCAGCCCG 2197
Db 1974 TGGAGGTGAACATCGTGACCGAGACCGAGTACGCGCTGGGCGATCATCCAGGCCAGGCCG 2033
Qy 2198 ACAAGAGCGAGGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGG 2257
Db 2034 ACAAGAGCGAGGCGAGCTGTGTGAGCCAGATCATCGAGCAGCTGATCAAGAGAGAGAGG 2093
Qy 2258 TGTAACCTGAGTGGGTGCCCGCCCAAGGGCATCGGCGGCAACGAGCAGATCGACAAGC 2317
Db 2094 TGTAACCTGGCTGGGTGCCCGCCCAAGGGCATCGGCGGCAACGAGCAGGTGGAACAAGC 2153
Qy 2318 TGGTACCAAGGGGATCGGCAGAGTGTCTTCTGGAGCGGATCGATCGGCGGATCTGGA 2377
Db 2154 TGGTACCGCGGATCGGCAGAGTGTCTTCTGAAACGCGCATCGATCGATGGCGGATCTGGA 2213
Qy 2378 TCTACAGTACATGAGCAGCTGTACGTGGGACGCGGCGCTAGGATCGATTAAAGC 2437
Db 2214 TCTACAGTACATGAGCAGCTGTACGTGGGACGCGGCGCTAGGATCGATTAAAGC 2273
```

```
Qy 2438 TTCCCGGGGTAGCACCGGTGAATTC 2463
Db 2274 TTCCCGGGGTAGCACCGGTGAATTC 2299

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-synGp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 67.1%; Score 1651.8; DB 4; Length 4307;
Best Local Similarity 82.5%; Pred. No. 4.2e-273;
Matches 1939; Conservative 0; Mismatches 391; Indels 19; Gaps 4;

Qy 33 GGCACACGAGCCCAACATCTGATGAGCGGAGCAACTTCAAGGGCCCCAAGCGCATCAT 92
Db 1110 GACCAACTCCGCTACCATCATGATGACGCGCGGCAACTTTTCGGAACCAACGCAAGATCGT 1169
Qy 93 CAAGTGTCTCAACTGCGCAAGGAGGCGCACATCGCCGCAACTGCGCGCCCCCGCAA 152
Db 1170 CAAGTGTCTCACTGTGCGCAAGGAGGCGCACAGCCGCGCAACTGCGAGGGCCCCTAGGA 1229
Qy 153 GAAGGGTGTCTGGAAGTGC CGGCAAGAGGCGCACACAGATGAAGACTGCACCGAGCGCA 212
Db 1230 AAAGGGTGTCTGGAATGCGCAAGGAGGCGCACAGATGAAGACTGTACGGAGAGACA 1289
Qy 213 GGCACACTTCTTCGCGAGGACCTGGCTTCCCCCAGGGGCAAGGCGCGAGTTCCCCAG 272
Db 1290 GGCTAA-TTTTTTAGGGAAGATCTGGCTTCTCTACAAGGGAAGGCGAGGAAATTTTCTC 1348
Qy 273 CGAGCAGAACCGCGCCAAACAGCCCCACAGCGCGAGCTGCAGGT-----GCGCGGCGA 326
Db 1349 AGAGCAGACCGCGCCCAACAGCCCCACCGCGCGAGCTTCAGGTCTGGGGTTCGGGACAA 1408
Qy 327 CAACCCCCGAGAGGAGCCGCGCGCGAGCGCGCAGGGCA-----CCCTGAATTTCCCCCA 380
Db 1409 CAATCCCCCTCCGAAGCAGGAGCGCGAGCGGCGACCGGTGTCTTCAACTTTCCCTCA 1468
Qy 381 GATCACCCTTGTGAGCGCGCCCCCTGGTGGAGCATCAAGTGGCGCGCCAGATCAAGAGGC 440
Db 1469 GGTACACCTTTGGCAGCGACCCCTCGTCAACATCAAGATCGGGGGGCGAGCTCAAGAGGC 1528
Qy 441 CCTGCTGACACCGCGCGCGACGACACCGTGTCTGGAGAGATGAGCTGCCCGGCAAGTGTG 500
Db 1529 TCTCTGGACACCGGAGACGACGACCGTGTCTGGAGAGATGTCGTGTCCAGGCGCGCTG 1588
Qy 501 GAAGCCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGCCAGTACGACCAAGAT 560
Db 1589 GAAGCCGAAGATGATCGGGGGAATCGCGGTTTTTCATCAAGGTGCGCCAGTATGACCAAT 1648
Qy 561 CTTGATCGAGATCTGCGGCAAGAGGCGCATCGGCACCGTGTCTGATCGGCCCCCAGCCCGGT 620
Db 1649 CCTCATCGAAATCTCGCGCCCAAGGCTATCGGTACCGTGTCTGTGGGGCCCCCAGCACCGGT 1708
Qy 621 GAACNTATCGGCGCAACATGCTGACCCAGCTGGGCTGGCCTGCACCTGAACTTCCCATCAG 680
Db 1709 CAACNTATCGGAGCAACCTGTTGAGCGAGATCGGTTGACAGCTGAACTTCCCATTTAG 1768
```





Matches 1927; Conservative 0; Mismatches 403; Indels 19; Gaps 4;			
Qy	33	GGCCACAGCGCCCAACATCTCTGATGCGCGCAGCAACTTCAAGGGCCCGCCCAAGCGCATCAT	92
Db	1110	GACCAATCCCGTACCATCATGATGATGAGCGCGGCAACTTTCGGAAACCAACGCAAGATCGT	1169
Qy	93	CAAGTGTCTTCACTGCGCGCAAGAGGGCCACATCGCCCGCAACTGCGCGCCCGCCCGCAA	152
Db	1170	CAAGTGTCTTCACTGTGCGCAAGAGGGCCACATCGCCCGCAACTGCGCGCCCGCCCGCAA	1229
Qy	153	GAAGGGTGTCTGGAAGTGGCGCAAGAGGGCCACCATGATGAAGACTGCAACCGAGCGCCA	212
Db	1230	AAAGGGTGTCTGGAAGTGGCGCAAGAGGGCCACCATGATGAAGACTGTAAGAGAGACA	1289
Qy	213	GGCCAACTTCTTCGCGAGGACCTGGCTTTCCCGCAGGCAAGGGCCCGCGAGTTTCCCGAG	272
Db	1290	GGCTAA-TTTTATAGGGAAGATCTGGCTTCTTACAGGGAAGGCGCAGGGAATTTTCTTC	1348
Qy	273	CGAGCAGAACCGCGCCCAACAGCCCCACAGCGCGAGCTGAGTGGCGG-----CGA	326
Db	1349	AGAGCAGACAGAGCCCAACAGCCCCACAGAGAGAGCTTCAAGTCTGTGGGTAGAGACAA	1408
Qy	327	CAACCCCGCAGGAGGCGCGCGCCGAGCGCCAGGGCA-----CCCTGAACTTTCCCCCA	380
Db	1409	CAACTCCCCCTCAGAAGCAGAGCCGATAGACAAGGAATGTATCTTTAACTTTCCCTCA	1468
Qy	381	GATCACCTCTGGCAGCGCCCGCTGTGTGAGCATCAAGTGGCGGCGCAGATCAAGGAGGC	440
Db	1469	GATCACTTWTGGCAACGACCCCTCGTCAATAAAGATAGGGGGGCGAGCTCAAGGAGGC	1528
Qy	441	CTTGCTGGACACCGCGCGCGAGCAGCACCGTGTGGAGGAGATGAGCCTGCGCGGCAAGTG	500
Db	1529	TCTCTTGGAACCGGAGCAGACACACCGTGTGGAGGAGATGTCTGTGCGAGCGCGCTG	1588
Qy	501	GAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGCAGTACGACAGAT	560
Db	1589	GAAGCCCAAGATGATCGCGGGAATCGCGGCTTCATCAAGGTGGCGCAGTATGACCCAGAT	1648
Qy	561	CTTGATCGAGATCTCGCGCAAGAGGCCATCGGCACCGTGTGATCGGCGCCCGCCCGCT	620
Db	1649	CCTCATCGAATCTGCGGCCACAGGCTATCGGTACCGTCTGTGTGGCGCCCGCACACCGT	1708
Qy	621	GAACATCATCGCGCGCAACATGTGTACCCAGTGGGTGCACTCTGAACTTTCCCGCATCAG	680
Db	1709	CAACATCATCGGACGCAACCTGTGTGACGAGATCGGTGTCACGCTGAACTTTCCCGCATAG	1768
Qy	681	CCCCATCGACACCGTGCCTGAGCTGAGCCCGGATGAGCGGCGCCCGCAAGGTGAGCA	740
Db	1769	CCCTATCGACACGCTACCGGTGAAGCTGAAGCCCGGATGAGCGGCGCCCGCAAGGTGAGCA	1828
Qy	741	GTGGCCCTTGACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGATGGAGAA	800
Db	1829	ATGGCCATTGACAGAGAGAGAGATCAAGGCATCTGGTGGAGTTTGCACAGAGATGNAAA	1888
Qy	801	GGAGGGCAAGATCACAAGATCGGCCCGCGAGAACCCCTCAACAACCCCGTGTTCGCGCAT	860
Db	1889	GGAAAGGAAAATCTCCAAGATTGGGCTGTGAAGACCCCGTCAACAACGCGCGTGTTCGCAAT	1948
Qy	861	CAAGAGAGAGCAGCAGCAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGCAACAGCG	920
Db	1949	CAAGAGAGAGCAGCAGCAGTGGCGCAAGCTGGTGGACTTTCGCGAGCTGCAACAGCG	2008
Qy	921	CACCCAGGACTTCTGGAGAGTGCAGCTGGGCATCCCCCGCCCGCGCTGGAAGAGAA	980
Db	2009	CACGCAAGACTTCTGGAGAGTTCAGCTGGGCGATCCCCCGCACCCCGCAGGGCTGGAAGAA	2068
Qy	981	GAAGAGCGTACCGTGTGACCGTGGCGCAGCGCTACTTCAGCGTGCCTCTGACGAGGA	1040
Db	2069	GAATCCGTGACCGTACTGATGTGGTGTGATGCTACTTCTCGGTTCCCTCGACGAGAA	2128
Qy	1041	CTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAAGAGACCCCGCGCATCG	1100
Db	2129	CTTCAGGAAGTACACTGCTCTCAATCCCTTCGATCAACAAGAGACACCGGGGATTCG	2188

Qy	1101	CTACCAAGTACAACGCTGCTGCCCGAGGGCTGGAAGGGCAGCGCCCGCAGCATCTTTCAGAGCAG	1160
Db	2189	ATATCAGTACAACGCTGCTGCCCGAGGGCTGGAAGGGCTCTCCGCAATCTTCCAGAGTAG	2248
Qy	1161	CATGACCAAGATCCTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCTACCA---	1217
Db	2249	CATGACCAAAATCCTGGAGCCTTTCCGCAAAACAGAACCCCGACATCTCTATCAGTA	2308
Qy	1218	---GGCCCCCTCTAGCTGGGCGAGCGACTGGAGATCGGCGCAGCACCGCGCAAGATCGA	1274
Db	2309	CATGAGATGACTTGTACGTGGGCTCTGATCTAGAGATAGGCGAGCACCGCACCAAGATCGA	2368
Qy	1275	GGAGCTGCGCAAGCACTGCTGCGCTGGGCTTCAACACCCCGACAAAGAACACCCAGAA	1334
Db	2369	GGAGCTGCGCGCAGCACCTGTTGAGTGGGACTGACCAACCCCGCAAGAACACCCAGAA	2428
Qy	1335	GGAGCCCCCTTCTCTGTGATGGCTACGAGCTGCAACCCCGACAAAGTGGACCGTGCAGCC	1394
Db	2429	GGAGCCCCCTTCTCTGTGATGGGTTACGAGCTGCAACCCCGACAAATGGAACCGTGCAGCC	2488
Qy	1395	CATCGAGCTGCCCGAGAGGAGAGCTGGACCGCTGAAACGACATCCAGAAAGCTGGTGGGCAA	1454
Db	2489	TATCGTGTGCGCAGAGAAAGACAGCTGGAATGTCAACGACATACAGAAAGCTGGTGGGAA	2548
Qy	1455	GCTGAATCTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCGAGCTGTGCAAGCTGCT	1514
Db	2549	GTTGAATCTGGGCCAGTCAGATTTACCCAGGGATTAAAGTGGAGCAGCTGTGCAAACTCCT	2608
Qy	1515	GCGCGGCGCAAGCCCTGACCGACATCTGTCGCCCTGACCGAGGAGGCGCAGCTGGAGCT	1574
Db	2609	CCGCGAAACCAAGGCACTCAAGAGGTGATCCCCCTAACCGAGGAGGCGCAGCTCGAACT	2668
Qy	1575	GGCGGAGAACCGGAGATCTCGCGAGCCCGTGCAGCGGCTGTACTACCAACCCCGACAA	1634
Db	2669	GGCAGAAACCGAGAGATCTTAAAGAGCCCGTGCACCGGCTGTACTATGACCCCTCAA	2728
Qy	1635	GGACCTGTGTGGCGCAGATCCAGAAAGCAGGGCCACGACCAAGTGGACCTTACAGATCTTACCA	1694
Db	2729	GGACCTGTATCGCGCAGATCCAGAAAGCAGGGGCAAGGCGCAGTGCACCTTATCAGATTTACCA	2788
Qy	1695	GGAGCCCTTCAAGAACCTGAGAGCCGCAAGTACGCCAAGATGGGCACCGCCCGCACACAA	1754
Db	2789	GGAGCCCTTCAAGAACCTGAGAACCGGCAAGTACGCCCGGATGAGGGGTGCCACACTAA	2848
Qy	1755	CGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCGCCATGGAGAGCATCGTGATCTG	1814
Db	2849	CGAGCTGAAGCAGCTGACCGAGGCGGTGCAAGAGATCACCCCGAAAGCATCTGTGATCTG	2908
Qy	1815	GGCAAGAGACCCCAAGTTCCGCTGCCCTCCATCCAGAAAGGAGACCTTGGGAGACCTTGTGGAC	1874
Db	2909	GGCAAGAGACTCTTAAGTTCAAGTGCCTCCATCCAGAGGAAACCTTGGGAAACCTTGTGGAC	2968
Qy	1875	CGACTATGCGCAGGCCCTCTGGATCCCGAGTGGGAGTTCTGTGAAACACCCCGCCCTTGGT	1934
Db	2969	AGAGTATTTGGCAGGCCACCTGGATCTCTGAGTGGGAGTTCTGTCAACACCCCTTCCCTGGT	3028
Qy	1935	GAAGCTGTGTACACGTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGA	1994
Db	3029	GAAGCTGTGTATCACGTGGAGAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGA	3088
Qy	1995	CGCGCGCGCCAAACCGGAGACCAAGATCGGCAAGCGCGGTACGTGACCGACCGGGGCGG	2054
Db	3089	TGGGCGCGCTTAACAGGAGACTAAGCTGGGCAAGCGGATACGTCTACTAACCGGGGCGG	3148
Qy	2055	GCAGAGATGCTGAGCGCTGACCGAGACCAACCAAGAGAGACCGAGCTGCGAGGCGCATCA	2114
Db	3149	ACAGAGGTTGTCACTCTCACTGACCAACCAACAGAGACTGAGCTGCGAGGCGCATTTA	3208
Qy	2115	GCTGGCCCTGCGAGACGCGGAGGAGTGAACATCTGTGACCGACAGACCGGAGTACGCCCT	2174
Db	3209	CCTCGCTTTGAGGACTCGGGCCTGGAGGTGAACATCTGTGACAGACTCTCAGTATGCCCT	3268



Db 2569 GTTGAATGGGCGAGTCAGATTATCCACAGGANTAAAGTGAGGCGAGCTGTGCAAACTCCT 2628  
Qy 1515 GCGGGGCGCCAAAGGCGCTGACCGACATCGTGCCTCCCTGACCGAGGAGGCGGAGCTGAGCT 1574  
Db 2629 CCGCGAACCAGGCACTCACAGAGGTGATCCCTTAACCGAGGAGGCGGAGCTGCACT 2688  
Qy 1575 GCGCGAGAACCGGAGAGTCTCGGGAGCGCGTGCACGCGGTGTACTAGACCCCGAGAA 1634  
Db 2689 GGCAGAAACCGAGAGATCCTAAAGAGCGCGTGCACGCGGTGTACTATGACCCCTCCAA 2748  
Qy 1635 GGACTGTGTGCGCGAGATCCAGAGAGCGGCGCACGACAGTGGACCTACAGATCTACCA 1694  
Db 2749 GGACTGATGCCGAGATCCAGAGAGCGGCGAGCGGCGGACCTATCAGATTTACCA 2808  
Qy 1695 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGGCCACACAA 1754  
Db 2809 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCGGATGAGGGGTGCCACACTAA 2868  
Qy 1755 CGAGTGAAGCAGCTGACCGAGGCGGTGCAGAAATCGCCATGAGAGCATCGTGTATCTG 1814  
Db 2869 CGAGCTCAAGCAGCTGACCGAGGCGGTGCAGAAATCACACCGAAAGCATCGTGTATCTG 2928  
Qy 1815 GGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGGAGACCTGGGAGACCTGTGTGAC 1874  
Db 2929 GGGAAAGACTCTTAAGTTCAAGCTGCCCATCCAGAGGAAACCTGGGAAACCTGTGTGAC 2988  
Qy 1875 CGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGGT 1934  
Db 2989 AGATATTGGCAGGCCACCTGGATTCTGTAGTGGAGTTCTGTCAACACCCCTCCCTGGT 3048  
Qy 1935 GAAGCTGTGTACAGCTGGAGAGAGGACCCATCATCGGCGCGGAGACCTTCTACGTGA 1994  
Db 3049 GAAGCTGTGTACAGCTGGAGAGAGGACCCATATAGTGGCGCGCGAAACCTTCTACGTGA 3108  
Qy 1995 CGGCGCGCCAAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGCGCG 2054  
Db 3109 TGGGCGCGCTAACAGGAGACTAAGCTGGGCAAGCGGATACGTCACTAACCGGGGCGAG 3168  
Qy 2055 GCAGAAGATCGTAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAAGGCGCATCCA 2114  
Db 3169 ACAGAAGTTGTACCTCTACTGACACCAACCAAGAGACTGAGCTGCGAGGCGCATTA 3228  
Qy 2115 GCTGGCCCTGAGAGACGCGGAGGAGGTGAACATCGTGAACCGACGCGAGTACGCCCT 2174  
Db 3229 CCTCGCTTGCAGGACTCGGCGCTGGAGGTGAACATCGTGAACAGACTCTCAGTATGCCCT 3288  
Qy 2175 GGGCATCATCAGSCCGCGACAGAGCGAGCGAGTGTGTGAACACAGATCATCGA 2234  
Db 3289 GGGCATCATTCAGCCCGACGACAGAGTGTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGT 3348  
Qy 2235 GCAGCTGATCAAGAGGAGAGGTGTATCTGAGCTGGGTGCGCGCCCAAGAGGCGATCGG 2294  
Db 3349 GCAGCTGATCAAGAGGAGAGGTGTATCTGCGCTGGGTACCGCGCCCAAGAGGATGG 3408  
Qy 2295 CGGCAACGAGCAGATCGCAAGCTGGTGAACAGGCGATCGCAAGGTGTGTGTGTGTGTGT 2354  
Db 3409 CGGCAATGAGCAGTCCAGAGCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3468  
Qy 2355 CGGCATCGA 2363  
Db 3469 TGGCATCGA 3477

## RESULT 10

US-09-936-572-13

; Sequence 13, Application US/09936572

; Patent No. 6783981

; GENERAL INFORMATION:

; APPLICANT: UDEN, MARK

; APPLICANT: MITROPHANOUS, KYRIACOS

; TITLE OF INVENTION: ANTI-VIRAL VECTORS

; FILE REFERENCE: 078863/0137

; CURRENT APPLICATION NUMBER: US/09/936,572

; CURRENT FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: GB 9906177.2  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4353  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: pSVNGP3-codon  
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence from  
; OTHER INFORMATION: the major splice donor  
US-09-936-572-13

Query Match 66.0%; Score 1624.6; DB 4; Length 4353;

Best Local Similarity 81.8%; Pred. No. 1.9e+268;

Matches 1922; Conservative 0; Mismatches 408; Indels 19; Gaps 4;

Qy 33 GGCCACACGAGCGCCAAACATCTCTGATGCGCGCAGCAAACTTCAAGGGCGCCCAAGCGCATCAT 92  
Db 1156 GACCACTCCGCTACCATCATGATGCGCGCGCAACTTTTCGGAACCAACGCAAGATCGT 1215  
Qy 93 CAAGTGTCTTCAACTGCGGCAAGAGGCGCCACATCGCCCGCAACTGCCCGCCCCCGCAA 152  
Db 1216 CAAGTGTCTTCAACTGTGCAAGAGGCGCACACAGCGCGCAACTGCAGGGCCCCCTAGGAA 1275  
Qy 153 GAAGGGTGTGGAAGTGCAGCAAGGCGCCACAGATGAAGACTGCACGAGCGCA 212  
Db 1276 AAAGGGTGTGGAAGTGTGGAAGAGGAGCACCAAAATGAAGATTTGTACTGAGAGACA 1335  
Qy 213 GGCCAACTTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCGCGCGAGTTCCCGAG 272  
Db 1336 GGCTAA-TTTTTAAAGGAAGATCTGGCTTCCCAAGAGGAAGCCAGGGAATTTTCTTC 1394  
Qy 273 CGAGCAGAACCGCGCCAAACAGCCCCACAGCGCGAGCTGCAGGTGCGCGG-----CGA 326  
Db 1395 AGAGCAGACCAAGAGCCAAACAGCCCCACAGAAAGAGAGCTTCAGGTTTGGGGAAGAGACA 1454  
Qy 327 CAACCCCGCAGCGAGCG 380  
Db 1455 CAATCCCTCTTCAGAAAGCAGGAGCGGATAGACAAGGAACTGTATCTTTAGCTTCCCTCA 1514  
Qy 381 GATCACCTGTGGCAGCGCCCCCTGGTGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCG 440  
Db 1515 GATCACCTTTTGGCAGCGACCCCTCGTCAAAATGAAGATAGGGGGGCGAGCTCAAGGAGCG 1574  
Qy 441 CCTGCTGGACACCGCGCGCGACGACACCGTGTCTGGAGAGATGAGCTTCCCGCGCAAGTG 500  
Db 1575 TCTCTGGACACCGGAGCAGACGACACCGTGTGGAGGAGATGCTGTTGCCAGGCGCGCTG 1634  
Qy 501 GAAGCCCAAGATGATCGGCGCATCGGCGGCTTTCATCAGGTGCGCGCGCGCGCGCGCGCG 560  
Db 1635 GAAGCCCAAGATGATCGGCGGAACTCGGCGGTTTCATCAAGGTGCGCGCGCGCGCGCG 1694  
Qy 561 CCTGATCAGATCTGCGCGCAAGAGGCGCATCGGCGACCGTGTGTGTGTGTGTGTGTGTGTGT 620  
Db 1695 CCTCATGAAATCTGCGCGCACAAAGGCTATCGGTACCGTGTGTGTGTGTGTGTGTGTGTGT 1754  
Qy 621 GAACATCATCGGCGCGCAACATGTGACCCAGCTGGGCTGCAACCTGTGAACCTTCCCGCAT 680  
Db 1755 CAACATCATCGGACGCAACCTGTTTGAACGAGATCGTGTGCAACGCTGAACCTTCCCGCAT 1814  
Qy 681 CCCCATCGAGACCGTGTGCGCGCATGAAGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCG 740  
Db 1815 CCCTATCGAGACGCTACCGGTGAAGCTGAAGCCCGCGGATGAGCGCGCGCGCGCGCGCGCG 1874  
Qy 741 GTGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGGAGGAGATGGAGAA 800  
Db 1875 ATGGCCATTGACAGAGAGAGATCAAGGGCACTGCTGGAGATTTGCAACAGAGATGGAGAA 1934

Qy	801	GGAGGGCAAGATCATCAAGATCGCCCGCAGAACCCCTTACAAACCCCGTGTTCGCAT	860
Db	1935	GGAAAGGAAATCTCTCAAGATTGGCCCTGAGAACCCGTACAAACGCGGTGTTCGAAT	1994
Qy	861	CAAGAAGAAGACAGCAACCAAGTCGCGCAAGCTGGTGACTTCGCCGAGCTGAACAGCG	920
Db	1995	CAAGAAGAAGACTTCGACGAATGGCGAAGCTGGTGACTTCGCCGAGCTGAACAGCG	2054
Qy	921	CACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCCGCTGAAGAAGAA	980
Db	2055	CACCGAAGACTTCTGGGAGGTTCAGCTGGGCATCCCGCACCCCGCAGGCTGAAGAAGAA	2114
Qy	981	GAAGAGCGTGACCTGTCTGGAGCTGGGCGACGCTACTTTCAGCGTGCCTTCGACGAGGA	1040
Db	2115	GAATTCGCTGACCGTACTGGATGTGGTGATGCTCTCTTCGCTCCCTCGACGAAGA	2174
Qy	1041	CTTCGGCAAGTACACCGCCTTCACCATCCCACGAGCATCAAACAAGAGACCCCGCATCCG	1100
Db	2175	CTTCAGGAAGTACACTCCCTTCACAATCCCTTCGATCAAACAAGAGACACCGGGGATTCG	2234
Qy	1101	CTACCACTCAACAAGTGTGCCCCAGGCTTGAAGGGCAGGCCACAGCATCTTCACAGACGAG	1160
Db	2235	ATATCAGTACAAAGTGTGCCCCAGGCTTGAAGGGCTCTCCGCGCATCTTCACAGTAG	2294
Qy	1161	CATGACCAAGATCTCGAGGCCCTTCGCGGCCGCAACCCCGAGATCGTGATCTACCA---	1217
Db	2295	CATGACCAAAATCTCGAGCCTTCCGCAAAACAGAACCCCGACATCGTCTATCATCAGTA	2354
Qy	1218	---GGCCCCCTGTACTGTGGGCAGCGACTTGGAGATCGGCGCAGCACCGCGCCAGATCGA	1274
Db	2355	CATGGATGACTTGTACTGTGGGCTCTGATCTAGAGTAGGGCAGCACCGCACCAAGATCTGA	2414
Qy	1275	GGAGCTCGGCAAGCACTGCTGCGCTGGGGCTTCACCACCCCGCAAGAAGACACACAGAA	1334
Db	2415	GGAGCTCGGCACAGCACTGTTGAGTGGGAGTCAACCAACCCGCAAGAAGACACACAGAA	2474
Qy	1335	GGAGCCCCCTTCTCTGTGGATGGGCTACGAGCTGCACCCCGCAGAGTGGACCGTGCAAGCC	1394
Db	2475	GGAGCCTCCCTTCTCTGGATGGGTTACGAGCTGCACCCCTGACAAATGGACCCGTGCAGCC	2534
Qy	1395	CATCGAGCTGCCGAGAAAGAGAGACTGGAACGTTGAACGCAATCCAGAAAGCTGGTGGGCAA	1454
Db	2535	TATCGTCTGCGACAGAAAGACAGCTGGACTGTCTACGCAATACAGAAAGCTGGTGGGGAA	2594
Qy	1455	GCTCAACTGGSCAGCCAGATCTACCCCGGCATCAAGGTGCGCGAGCTGTGCAAGCTGCT	1514
Db	2595	GTTGACTGGGCCAGTCAAGTTTACCAGAGATTAAGTGAAGGAGCTGTGCAAACTCCT	2654
Qy	1515	GCAGCGGCGCAAGCCCTGCACCGACATCTGTCGCCCTGACCGAGAGGCGCGAGCTGGAGCT	1574
Db	2655	CCGCGAAACCAAGCACTCAAGAGGTGATCCCCCTTAACCGAGAGGCCGAGCTCGAACT	2714
Qy	1575	GGCCGAGAAACCGCAGATCTCTGCGCAGACCGGTGCAACGGCGTGTACTACACCCACAGCAA	1634
Db	2715	GGCAGAAAAACGAGAGATCTCTAAAGGAGCCCGTGCACGGCGTGTACTATGACCCCTCCAA	2774
Qy	1635	GGACCTGTTGGCCGAGATCCAGAGCAGGGGCCAGCAGCTGGACCTTACAGATCTTACCA	1694
Db	2775	GGACCTGATCGCCGAGATCCAGAAAGAGGGGCAAGGCCAGTGGACCTTACAGATTTACCA	2834
Qy	1695	GGAGCCCTTCAAGAACTCTGAAGACCGGCAGTACGCCAAGATGCGCACCCGCCACACCAA	1754
Db	2835	GGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACTAA	2894
Qy	1755	CGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAGATCGCCATGGAGAGCATTCGTGATCTG	1814
Db	2895	CGAGCTCAAGCAGCTGACCGAGGCGGTGCAGAGATCAACACCGAAAGCATTCGTGATCTG	2954
Qy	1815	GGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTTGGAGACCTTGGTGAC	1874
Db	2955	GGGAAAGACTCTTAAGTTCGAAGCTGCCCATCCAGAAAGAACTTGGGAAACCTTGGTGGAC	3014
Qy	1875	CGACTACTGCGAGGCCACCTGGATTCGCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGT	1934

3015	DB	AGAGTATTGGAGGCCACCTTGGGATTCCTGAGTGGGAGTTCTGTAACACCCCTCCCTCGGT	3074
1935	QY	GAAGCTCTGGTACACAGCTGGAGAGGAGGCCATCATCGGCGCGAGACCTTCTACGTGGGA	1994
3075	DB	GAAGCTCTGGTACCAAGCTGGAGAGGAGCCCATAGTGGGCGCCGAAACCTTCTACGTGGGA	3134
1995	QY	CGGCGCCGCAACCGCGAGACCAAGATTCGGCAAGCGCGGTACGTGTACCGAACCGGGGCCG	2054
3135	DB	TGGGGCCGCTAAACAGGGAGACTAAGCTGGGCAAGCGGATACGTCACTAACCGGGGCAG	3194
2055	QY	GCAGGAAGATCGTGAGCTGTACCGAGACCAACCAACAGAGACCGAGCTGCAGAGCCATCCA	2114
3195	DB	ACAGAAGGTTGTCAACCCCTCACTGACCAACCAACCAAGAACTGTAGCTGCAGGCAATTTA	3254
2115	QY	GCTTGGCCCTTCGAGGACAGCGGCACGAGGTGAAACATCGTACCGCACGAGCCAGTACGCGCT	2174
3255	DB	CCTCGCTTTGAGGACTCGGGCTGTGGAGGTGAACATCGTGACAGACTCTCAGTATGCGCT	3314
2175	QY	GGGCATCATCCAGGCCACGCCGACAAAGACGAGAGCGAGCTGGTGAAACAGATCATCGGA	2234
3315	DB	GGGCATCATTTCAAGCCCGCCAGACCCAGAGTGTAGTCCGAGCTGGTCAATCAGATCATCGGA	3374
2235	QY	GCAGCTGATCAAGAAAGGAGGTGTACTGTAGCTGGGTGCCGCCACCAAGGGCATCGG	2294
3375	DB	GCAGCTGATCAAGAAGGAAAAGGTCTATCTGGGCTTGGGTATCCCGGCCCAAAAGGCATTTG	3434
2295	QY	CGGCAACGACGAGATCGACAAGCTGGTGTAGCAAGGGGCATCCGCAAGGTGCTGTTTCTTGGA	2354
3435	DB	CGGCATGACGAGGTTCACAAAGCTGGTCTCGGGCTGGCATCAGGAGGTGCTATTCCTTGGA	3494
2355	QY	CGGCATCGA	2363
3495	DB	TGGCATCGA	3503

```

RESULT 11
US-09-936-572-12
; Sequence 12, Application US/09936572
; Patent No. 6783981
; GENERAL INFORMATION:
; APPLICANT: UDEN, MARK
; APPLICANT: MITROPHANOUS, KYRIACOS
; TITLE OF INVENTION: ANTI-VIRAL VECTORS
; FILE REFERENCE: 078883/0137
; CURRENT APPLICATION NUMBER: US/09/936.572
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: PCT/GB00/01002
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: GB 9906177.2
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 4542
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP2-codon
; OTHER INFORMATION: optimised HIV-1 gagpol with leader sequence
US-09-936-572-12

```

	Query Match	66.0%	Score 1624.6	DB 4	Length 4642
	Best Local Similarity	81.8%	Pred. No. 1.9e-288		
	Matches 1922	Conservative	0	Mismatches 408	Indels 19
	Gaps				
Qy	33	GGCCACAGCGCCCAACATCTCTGATGCGAGCGCAGCAACTTCAAGGGGCCCCAAGCGCATCAT	92		
Db	1445	GACCACTCGCTACCATCATCATGATGCGAGCGCGCACTTTCGGAACCAACGCAAGATCGT	1504		
Qy	93	CAAGTGTCTTCAACTGCGGCAAGAGGGGCCACATCGCCCGCAAATGCGCGCCCGCCCGCAA	152		
Db	1505	CAAGTGTCTTCAACTGTGTGCAAGAGGGCACAGCCCGCACTGTGAGGSCCCTAGGAA	1564		



```

Db 3724 CGCAATGAGCAGGTGCGAAGCTGCTCTGGCTGCGCATCAGGAAGTGTCTATTCCTGGA 3783
QY 2355 CGGCATCGA 2363
Db 3784 TGGCATCGA 3792

RESULT 12
US-09-552-950-5
; Sequence 5, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552,950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 9772
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:psvNGP
US-09-552-950-5

Query Match 66.0%; Score 1624.6; DB 4; Length 9772;
Best Local Similarity 81.8%; Pred. No. 2.1e-268;
Matches 1922; Conservative 0; Mismatches 408; Indels 19; Gaps 4;

QY 33 GGCACACGCGCAACATCTGTATGAGCGCGCAGCACTTCAAGGGCCCAAGCGCATCAT 92
Db 2217 GAACCAACTCCGTACCATCATGATGAGCGCGCAACTTTTCGGAACAAGCAAGATCGT 2276
QY 93 CAAGTCTTCAACTGCGGCAAGAGAGGCGCACATCGCCCGCAACTGCGGCGCCCGCGCAA 152
Db 2277 CAAGTCTTCAACTGCGGCAAGAGAGGCGCACAGCCCGCAACTGCGGCGCCCGTAGGAA 2336
QY 153 GAAGGCGTCTGGAAGTGGCGCAAGAGGCGCACAGATGAAGGACTGCACCGAGGCGCA 212
Db 2337 AAAGGCGTGTGGAAATGTGGAAAGGAAGGACACCAAAATGAAAGATTGACTGAGAGACA 2396
QY 213 GGCCTAACTTCTCCGCGAGGACCTGCGCTTCCCGCAGGCGCAAGCGCGGAGTTCCCGAG 272
Db 2397 GGGTAA-TTTTTTTGAAGGAGATCTGGCTTCCCAAGGAAGGCGCAGGGAATTTTCTTC 2455
QY 273 CGAGCAGAACCGCGCCCAACAGCCCGCCACCGCGAGCTGCAGGTGCGCGG-----CGA 326
Db 2456 AGAGCAGACAGAGCCCAACAGCCCGCCACAGAGAGGCTTCAGGTTTGGGGAAGAGACAA 2515
QY 327 CAAACCCCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
Db 2516 CAACTCCCTCTCAGAAAGCAGGAGCCGATAGACAAAGGAATGTATCTTTAGCTTCCCTCA 2575
QY 381 GATCACCCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
Db 2576 GATCCTCTTTGGCAGGACCGCTCTGTCAATAAAGATAGGGGGGCGAGCTCAAGAGGC 2635
QY 441 CTTGCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
Db 2636 TCTCTGGACACCGGAGCAGACACCGCTGTGGAGGAGATGTCTGTGCGAGGCGCTG 2695
QY 501 GAAGCCCAAGATGATCGCGGCGATTCGGCGGCTTCATCAAGGTGCGCGCGCGCGCGCGCG 560
Db 2696 GAAGCCCAAGATGATCGCGGGAATTCGGCGGTTTTCATCAAGGTGCGCGCGCGCGCGCG 2755
QY 561 CTTGATCGAGATCTCGCGGCAAGAGGCGCATCGCACCGCTGCTCATCGGCGCGCGCGCGCG 620
Db 2756 CTTTCATCGAAATCTCGCGGCGCAAGGCTATCGGTATCGGTGCGGCGCGCGCGCGCGCG 2815
QY 621 GAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCGCATCAG 680

```

```

Db 2816 CAAACATCATCGGACGCAACTGTTGACGAGATCGGTTGACGCTGAACTTCCCGCATTAG 2875
QY 681 CCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGCGCCCGCAAGGTGAAGCA 740
Db 2876 CCGTATCGAGACGGTATCCCGTGAAGCTGAAGCCCGGATGAGCGCGCGCAAGGTCAAGCA 2935
QY 741 GTGGCCCTTGACCGAGGAGAAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGATGAGAA 800
Db 2936 ATGGCCATTTGACAGAGAGAGAGATCAAGGCACCTGTTGGAGATTTGACAGAGATGAAA 2995
QY 801 GGAGGCAAGATCAACAAGATGGGCCCCGAGAAACCCCTACAAACACCCCGTGTTCGCCAT 860
Db 2996 GGAAGGGAATACTCCAAAGATTGGGCTTGAGAACCCGTACAAACACCCCGTGTTCGCAAT 3055
QY 861 CAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAAGCG 920
Db 3056 CAAGAAGAAGGACTCGACGAAATGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAAGCG 3115
QY 921 CACCCAGGACTTCTGGGAGGTGAGCTGGGCACTCCCGCACCCCGCGCGCTGAAGAAGAA 980
Db 3116 CAGCGAAGACTTCTGGGAGGTTTCAGCTGGGCACTCCCGCACCCCGCAGGGCTGAAGAAGAA 3175
QY 981 GAAGAGCTGACCGTCTGGAAGCGTGGGCGACCGCTACTTTCAGGCTGCCCTTGAACGAGGA 1040
Db 3176 GAAATCCGTGACCGTACTGGAATGGGTGATGCTACTTCTCCGTTCCCGTGGACGAGGA 3235
QY 1041 CTTCCGCAAGTACACCGCTTCCACCATCCCGACATCAACAAGAGAGACCCCGGCATCCG 1100
Db 3236 CTTCCGCAAGTACACCGCTTCCACCATCCCGACATCAACAAGAGAGACCCCGGCATTCG 3295
QY 1101 CTACCAAGTACAACTGCTGCCCGCAGGGCTGGAGGGCAGCCCGCAGCATCTTTCAGAGCAG 1160
Db 3296 ATATCAGTACAGCTGCTGCCCGCAGGGCTGGAAAGGCTCTCCCGCAATCTTCCAGAGTAG 3355
QY 1161 CATGACCAAGATCTGGAAGCCCTTCCGCGCCGCAACCCCGAGATCTGTGATCTTACCA--- 1217
Db 3356 CATGACCAAAATCCTGGAGCCTTTCGCGCAACAGAAACCCCGCATCTCTATCAGTA 3415
QY 1218 ---GGCCCCCTGTAGCTGGGACGACCTGGAGATCGGCGACGCGCGCCCAAGATCGA 1274
Db 3416 CATGGATGACTTGTAGCTGGGCTCTGATCTAGAGATAGGGCAGACCGCACCAAGATCGA 3475
QY 1275 GGAGCTGCGCAAGCACCTGCTGCGCTGGGCTTTCACACCCCGCACAAGAAGACCAAGAA 1334
Db 3476 GGAGCTGCGCCAGCACCTGTTGAGGTGGGACTGACACACCCGACAAGAAGACCAAGAA 3535
QY 1335 GGAGCCCCCTTCTCTGTGATGGGCTACGAGCTGCAACCCGACAAGTGGACCGTGCAGCC 1394
Db 3536 GGAGCCCTCCCTTCTCTGTGATGGGTTACGAGCTGCAACCCCTGACAAATGGACCGTGCAGCC 3595
QY 1395 CATCGAGCTGCGCGAGAAGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGGTGGGCAA 1454
Db 3596 TATCGTGTGCGCAGAGAAAGACAGCTGGACTGTCAACGACATACAGAGCTGGTGGGAA 3655
QY 1455 GCTGAACCTGGGCGCAGCAGATCTTACCCCGCATCAAGGTGCGGCTGTGCAAGCTGCT 1514
Db 3656 GTTGAACCTGGGCGCAGTCTGATTTTACCCAGGATTAAGGTGAGGCTGTGCAAACTCCT 3715
QY 1515 GCGCGGCGCAAGGCCCTGACCGACATCTGTGCCCTTGAACCGAGAGGCGCGAGCTGGAGCT 1574
Db 3716 CCGCGGAAACCAAGGCACTCAGAGGTGATCCCGCTTAACCGAGGAGGCGCGAGCTCGAACT 3775
QY 1575 GCGCGGAGAACCGGAGATCTCTGCGGAGCCCGTGCAGCGGCTGTACTACGACCCAGCAA 1634
Db 3776 GGCAGAAACCGAGAGATCTTAAGGAGCCCGTGCAGCGGCTGTACTATGACCCCTCAA 3835
QY 1635 GGAACCTGGTGGCGCAGATCCAGAAAGCAGGCGCACGACCAAGTGGACCTTACAGATCTTACCA 1694
Db 3836 GGAACCTGATCGCGCAGATCCAGAAAGCAGGCGCAAGCCAGTGGACCTTATCAGATTTACCA 3895
QY 1695 GGAGCCCTTCAAGAACTGTGAAGACCGGCAAGTGAAGCAAGTGGCACCGCCACACCAA 1754

```



Db 3896 GGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCGGATGAGGGGTGCCCACTAA 3955  
Qy 1755 CGAGTGAAGCAGCTGACCGAGGCGGTGCAGAAGATCGCCATGGAGAGCATCGTGTATCG 1814  
Db 3956 CGAGCTCAAGCAGCTGACCGAGGCGGTGCAGAAGATCACACCGAAAGCATCGTGTATCG 4015  
Qy 1815 GGGCAAGACCCCAAGTTCGCTCCCTGCCATCCAGAAGAGACCTGGAGACTGGTGAC 1874  
Db 4016 GGGAAAGACTCTTAAGTTCAAGCTGCCATCCAGAAGAAACCTGGGAAACCTGGTGAC 4075  
Qy 1875 CGACTACTGGCAGGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCCCTGGT 1934  
Db 4076 AGAGTATGGCAGGCCACTGGATCTCTGAGTGGGAGTTCGTCAACACCCCTCCCTGGT 4135  
Qy 1935 GAAGCTGTGTPACAGCTGGAGAGAGGCCCATCATCGCCCGCAGACCTTCTACGTGA 1994  
Db 4136 GAAGCTGTGTPACAGCTGGAGAGAGGCCCATAGTGGGCGCGGAACCTTCTACGTGA 4195  
Qy 1995 CGGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGCG 2054  
Db 4196 TGGGCGCGCTTAACAGGGAGACTAAGCTGGGCAAGCGGATACGTCAACCGGGCGAG 4255  
Qy 2055 GCAGAAGATCTGAGCTGACCGAGACGAGACCAACCAACAGAGACCGAGCTGCAGGCCATCCA 2114  
Db 4256 ACAGAAGTGTACCTCTACCTCACTGACACCAACCAAGAGACTGAGCTGCAGGCCATTTA 4315  
Qy 2115 GCTGGCCCTGAGGACAGCGCAGCGAGGTGAACATCGTACCGACAGCCAGTACGCCCT 2174  
Db 4316 CCTCGCTTTCAGGACTCGGSCCTGGAGGTGAACATCGTGCAGACTCTCAGTATGCCCT 4375  
Qy 2175 GGGCATCATCAGGCCAGCCCGCAAGAGCGAGCGAGCTGTGTGAACAGATCATCGA 2234  
Db 4376 GGGCATCATCAGGCCAGCCAGCAGAGTGAGTCCGAGCTGTGTGAACAGATCATCGA 4435  
Qy 2235 GCAGCTCATCAAGAGAGAGGTGTACCTGAGCTGGTGGCGCCCGCCACAGGCGCATCGG 2294  
Db 4436 GCAGCTCATCAAGAGAGAAAGGTGTATCTGGCTGGGTACCCCGCCACAAAGGCAATGG 4495  
Qy 2295 CGGCAACGAGCAGATCGACAGCTGGTGAGCAAGGGCATCCGCAAGGTGTGTTCCTGGA 2354  
Db 4496 CGGCAATGAGCAGTTCGACAGCTGGTCTCGGCTGGCATCAGGAAGGTGTATTCCTGGA 4555  
Qy 2355 CGGCATCGA 2363  
Db 4556 TGGCATCGA 4564

## RESULT 13

US-09-872-733A-6  
; Sequence 6, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, HIV GAG AND  
; TITLE OF INVENTION: HIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, HIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 8366  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
; OTHER INFORMATION: of the construct pCMVgagpolENkan containing a CMV  
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin  
; OTHER INFORMATION: resistance gene

## US-09-872-733A-6

Query Match 63.6%; Score 1566; DB 4; Length 8366;  
Best Local Similarity 80.2%; Pred. No. 2e-258;  
Matches 1885; Conservative 0; Mismatches 446; Indels 19; Gaps 4;  
Qy 33 GGCACACAGCGCCAAATCTCTGATGCGAGCGCAACATTTCAAGGGCCCCCAAGCGCATCAT 92  
Db 1879 GACGAATCTCGCGCACCAATATGATGCGAGAGGCAACTTCCGGAAACCAAGCGGAAGATCGT 1938  
Qy 93 CAAGTGTCTCAACTGCGCGCAAGAGGCGCCACATCGCCGCAACTGCGCGCCCCCGCAA 152  
Db 1939 CAAGTGTCTCAATTTGTGCAAGAGGCGCACCGCCAGGAACTGCGGGCCCCCGGAA 1998  
Qy 153 GAAGGGTCTGTGAAGTGCAGCAAGAGGGGCCACACAGATGAAGGACTGCACCGAGCGCA 212  
Db 1999 GAAGGGTCTGTGGAATGTGGAAGAGGAGACCAAAATGAAGATTGTACTGAGAGACA 2058  
Qy 213 GGCCAACTTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGGCGCGGAGTTCGCCAG 272  
Db 2059 GGCTAA-TTTTTTAGGGAAGATCTGGCTTCTCTAAGGGGAAGGCCAGGGAATTTTCTTC 2117  
Qy 273 CGAGCAGAACCGCGCCCAACAGCCCCACAGCGCGAGCTGCAGGTGCGCG- ---CGA 326  
Db 2118 AGAGCAGACAGAGCCCAACAGCCCCACAGAGAGAGCTTCAGGTCTGGGGTAGAGCAA 2177  
Qy 327 CAACCCCGCAGAGGCGCGCGCGAGCGCCAGGGCA- ---CCCTGAATTTCCCCCA 380  
Db 2178 CAATCCCTCCAGAGCAGGAGCGGATAGACAAGAACTGTATCTTAACTTCCCTCA 2237  
Qy 381 GATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGTGGGCGGCCAGATCAAGAGGC 440  
Db 2238 GATCACTCTTTGGCAACGACCCCTCGTCAAGTAAGGATCGGGGGCAACTCAAGGAAGC 2297  
Qy 441 CCTGCTGGAACCGCGCGCGAGACCGTCTGAGGAGATGACCTTGGCCCCCAAGTG 500  
Db 2298 GCTGCTCGATACAGGAGCAGATGATACAGTATTAGAAGAAATGATTTGCCAGGAAGATG 2357  
Qy 501 GAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCCAGTACGACCCAGAT 560  
Db 2358 GAACCAAAATGATAGGGGGGATCGGGGGCTTCATCAAGGTGAGGCGATGACGACCAT 2417  
Qy 561 CTGATCGAGATCTGGCGCAAGAGGCGCATCGGCACCGTGTCTGATCGGCCCCACCCCGT 620  
Db 2418 ACTCATAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTCACCTGT 2477  
Qy 621 GAACATCATCGCGCAACATGCTGACCCAGCTGGGCTGCACCTCTGAACCTTCCCCATCAG 680  
Db 2478 CAACATAATTGGAAGAAATCTGTTGACCCAGATCGGCTGCACCTTGAACCTTCCCATCAG 2537  
Qy 681 CCCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGCATGGACGCGGCCCAAGGTGAAGCA 740  
Db 2538 CCTATTGAGACGGTGCCTGGAAGTTGAAGCCGGGATGGACGCGGCCCAAGGTCAAGCA 2597  
Qy 741 GTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGGAGAA 800  
Db 2598 ATGGCCATTGACGAAAGAGAGATCAAGGCGCTTAGTCCAAATCTGTACAGAGATGGAGAA 2657  
Qy 801 GGAGGGCAAGATCACCAAGATCGCCCCCGAGAACCCCTACAAACCCCGCTGTTCGCCAT 860  
Db 2658 GGAAGGGGAAGATCAGCAAGATCGGGCTTGAGAACCCCTACAAACACTCCAGTCTTCGAAT 2717  
Qy 861 CAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGGAACAAGCG 920  
Db 2718 CAAGAAGAAGACAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAGCTGAACAAGAG 2777  
Qy 921 CACCCAGGACTTCTGGAGAGTGCAGCTGGGCATCCCCCAACCCCGCGCTGGAAGAGAA 980  
Db 2778 AACTCAGGACTTCTGGGAAGTTCAGCTGGGCATCCACATCCCGCTGGGTGGAAGAA 2837  
Qy 981 GAAGAGGTGACCGTGTGACAGTGGCGAGCGCTACTTACAGGTGCCCCCTGGACGAGGA 1040  
Db 2838 GAAGTCAAGTACAGTGTGAGTGGGTGATGCTTCTTCTCCGTTCCTTGGACGAGGA 2897

```
QY 1041 CTTCCGCAAGTACACCGCTTACCATCCACGATCAACAACGAGAGACCCCGGCATCG 1100
Db |||||
QY 2898 CTTTCAGAGAGTACCTGCTCCCTTACGATACCTAGCATCAACACGAGACACCGGCATCG 2957
Db |||||
QY 1101 CTAACAGTACAACGCTGCTGCCAGGGCTGGAAGGGCAGCCCGAGCATCTTCCAGAGCAG 1160
Db |||||
QY 2958 CTAACAGTACAACGCTGCTGCCAGGGATGGAAGGGATCACAGCCATCTTTCAAAGCAG 3017
QY 1161 CATGACCAAGATCCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCTACCA--- 1217
Db |||||
QY 3018 CATGACCAAGATCCTGGAGCCCTTCGCGCAAGCAAAAACCCAGACATCGTGATCTATCAGTA 3077
QY 1218 ---GGCCCCCTCTAGCTGGGAGGACCTTGAGATCGGCCAGCACCGCGCCCAAGATCGA 1274
Db |||||
QY 3078 CATGAGACGACCTCTAGTGAAGAGTGAACCTTGGAGATCGGGCAGCACAGACCAAGATCGA 3137
QY 1275 GGAGCTGCGCAAGCACCTGTGCGCTGGGGCTTTCACACCCCCCGACAAGAAGCACCAAGAA 1334
Db |||||
QY 3138 GGAGCTGAGACAGCATCTGTTGAGGTGGGACTGACCACACAGACAAGAAGCACCAAGAA 3197
QY 1335 GGAGCCCCCTTCTGTGTGATGGGCTACGAGCTGCAACCGGACCAAGTGACCGTGAGCC 1394
Db |||||
QY 3198 GGAACCTCCCTTCTGTGTGATGGGCTACGAACTGCATCTCTGACAAGTGAGACAGTGCAAGC 3257
QY 1395 CATCGAGCTGCCGAGAGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGCTGGGCAA 1454
Db |||||
QY 3258 CATCGTGTGCTTGAAGAGACAGCTGGACTGTGAACGACATACAGAGCTCTGGGGCAA 3317
QY 1455 GCTGAACCTGGGCGAGCAGATCTACCCCGGCATCAAGGTGCGCAGCTGTGCAAGCTGCT 1514
Db |||||
QY 3318 GTTGAACCTGGCAAGCAGATCTACCCAGGACATCAAAAGTTAGGACGCTGTGCAAGCTGCT 3377
QY 1515 GCGCGGGCCAGGCCCCGTGACCGACATCTGTGCCCTTGACCGAGGAGGCGCGAGCTGAGCT 1574
Db |||||
QY 3378 TCAGAGAAACCAAGGCACTGACAGAAGTGATCCCACTGACAGAGGAAGCAGAGCTAGAAT 3437
QY 1575 GGCAGGAGAACCGAGATCTGCGCAGCCGTCGACGCGCTGTACTACGACCCAGCAA 1634
Db |||||
QY 3438 GGCAAGAACCGAGATCTGGAAGAGCAGTACATGGAGTGATACGACCCAGCAA 3497
QY 1635 GGACCTGTTGGCGCAGATCCAGAAGCAGGGCCACGACCAAGTGACCTACAGATCTACCA 1694
Db |||||
QY 3498 GGACCTGATCGCAGAGATCCAGAAGCAGGGGCAAGCCCAATGACCTACCAATCTACCA 3557
QY 1695 GGAGCCCTTCAAGAACTGGAAGACCGGCAAGTAGGCAAGATCGGCACCGCCACACCAA 1754
Db |||||
QY 3558 GGAGCCCTTCAAGAACTGGAAGACAGGCAAGTAGGCAAGATGAGGGGTGCCCAACCAA 3617
QY 1755 CGACGTGAAGCAGCTGACGAGGCGCTGCAGAAGATCGCCATGGAGGATCGTGATCTG 1814
Db |||||
QY 3618 CGATGTGAAGCAGCTGACAGAGGAGTGCNAGAAGTACCAACAGAGAGATCTGTATCTG 3677
QY 1815 GGGCAAGACCCCAAGTTCCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGGAC 1874
Db |||||
QY 3678 GGGCAAGATCTCCAAGTTCAAGTGGCCATACAGAAGGAGACATGGAGACATGGTGGAC 3737
QY 1875 CGATCTGTCAGGCACTCTGATTCGCCAGTGGGAGTTGCTGAACAACCCCCCTCTGCT 1934
Db |||||
QY 3738 CGAGTACTGCAAGCCACCTGGATCTGAGTGGGAGTTGCTGAACAACCCCTCTCTGCT 3797
QY 1935 GAAGCTGTGTTACAGCTGGAGAAGGAGCCCATCATCGCGCCCGAGACCTTCTACGTGGA 1994
Db |||||
QY 3798 GAAACTGTGTTATCAGCTGGAGAAGGAACCCATCTGTGGGAGCAGAGACCTTCTACGTGGA 3857
QY 1995 CGCGCGCGCCAAACCGGAGACCAAGATCGGCAAGGCGCGGTACGTGACCGACCGGGGCGG 2054
Db |||||
QY 3858 TGGGGCAGCCAAACAGGGAGACCAAGCTGGGCAAGGCGAGCTAGTGAACCAACCGAGAGC 3917
QY 2055 GCAGAAGATGCTGAGCTGACCGAGACCAACCAACAGAGACCGAGCTGACGGCCATCCA 2114
Db |||||
QY 3918 ACAGAAGATGTTGACCTGTACACCAACCAACAGAGACTGAGCTGCAAGCCATCTA 3977
```

```
QY 2115 GTTGCCCTGCGAGGACAGCGCAGGAGGTGAACATCTGTCACCGACAGCCAGTACGCCCT 2174
Db |||||
QY 3978 CTTAGCTCTGCAAGACAGCGGACTGGAAGCTGAACATCTGTCAGACAGTCTACAGTACGCACT 4037
QY 2175 GGGCATCTCCAGGCCCGCAGCCGACAAGAGCAGAGCAGAGCTGGTGAACAGATCATCGA 2234
Db |||||
QY 4038 GGGCATCTCCAGCACAACCAAGCAATCCGAGTCAGAGCTGGTGAACAGATCATCGA 4097
QY 2235 GCAGCTGTACAAGAGGAAGGTGTACCTGAGCTGGGTGCCGCCACAAGGGCATCGG 2294
Db |||||
QY 4098 GCAGCTGTACAAGAGGAAGGTGTACCTGGATGGGTACCAGCACACAAGGAATTGG 4157
QY 2295 CGGCAACGAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCTGGA 2354
Db |||||
QY 4158 AGGAATGAACAGTAGATAAATTAGTCAGTGTGGATCCGGAAGGTGCTGTCTCTGGA 4217
QY 2355 CGGCATCGAT 2364
Db |||||
QY 4218 CGGGATCGAT 4227

RESULT 14
US-09-872-733A-1
; Sequence 1, Application US/09872733A
; Patent No. 6656706
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America, as
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND
; TITLE OF INVENTION: SIV ENV GENES
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV
; CURRENT APPLICATION NUMBER: US/09/872,733A
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US00/34985
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/173,036
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Mutated Human
; OTHER INFORMATION: Immunodeficiency Virus - 1 Gag/Pol gene
US-09-872-733A-1

Query Match 62.1%; Score 1530.6; DB 4; Length 4338;
Best Local Similarity 80.1%; Pred. No. 2e-252;
Matches 1905; Conservative 0; Mismatches 444; Indels 28; Gaps 8;

QY 14 TGGCCGAGGCCATGAGCCAGGCCACCAGC---GCCAACAATCTCTGATGTCAGCGCAGCAACT 70
Db 1085 TGGCCGAGGCCATGAGCCAGGCCAGGTGACGAATCTCGCGACCAATAATGATGTCAGAGAGGCAACT 1144
QY 71 TCAAGGGCCCCAAGCGCATCATCAAGTGTCTTAACTGCGCAAGAGAGGGCCACATCGCCC 130
Db 1145 TCGGAACACAGCGGAAGATCGTCAAGTGTCTTAAATTGTGGCAAGAGGGCACACCGCCA 1204
QY 131 GCAACTGCGCGCCCCCGCAGGCCAGGCTGTGGAAGTGTGGAAGTGTGGAAGAGGGCCACAGA 190
Db 1205 GGAACCTGCGGGGCCCCCGAAGAGGGCTGTTGGAATAATGTGGAAGAGGAGCACCAAA 1264
QY 191 TGAAGACTGTCACCGAGCCAGGCCAATCTTCTTCGCGAGGACCTGGCTTCCCCCAGG 250
Db 1265 TGAAGATTTGTACTGAGACAGAGGCTAA-TTTTATTAGGGAAGATCTGGGCTTCTTCAAG 1323
QY 251 GCAAGGCCCGCAGTTTCCCGAGCGAGCAAAACCGCCCAACAGCCCAACAGCGCGGAGC 310
Db 1324 GGAAGGCCAGGGNAATTTTCTTCAGACAGACACAGAGCCCAACAGCCCAACAGAGAGC 1383
QY 311 TGCAGTGTGCGG-----CGACAACCCCGCAGCGAGCGCGCGCGCAGCGCA 364
Db |||||
```

Db 1384 TTCAAGTCTGGGTAGAGACAACAACTCTCCCTCAGAAAGAGAGCCGATAGACAAGGAA 1443  
Qy 365 -----CCCTGAACCTTCCCCAGATCACTGTGTGGCAGCGCCCCCTGTGTAGCATCAAGG 418  
Db 1444 CTGTATCTTTAACTTCTCCCTCAGATCACTCTTTGGCAACGACCCCTGTGTACAGTAAGGA 1503  
Qy 419 TGGCGGCCAGATCAAGGAGCCCTGTGTGGACACCGGCGCGAGACACCGTGTGGAGG 478  
Db 1504 TCGGGGGGCAACTCAAGGAAGCGTGTGTGATACAGGAGCAGATGATACAGTATTAGAAG 1563  
Qy 479 AGATGAGCCTCCCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCA 538  
Db 1564 AAATGAGTTTCCAGGAAGATGGAAACCAAAATGATAGGGGGATCGGGGGCTTCATCA 1623  
Qy 539 AGTGCGCCAGTACGACACAGATCTCTGATCGAGATCTGCGGCAAGAAGGCCATCGGCACCG 598  
Db 1624 AGGTGAGCGAGTACGACCCAGATCTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 1683  
Qy 599 TGCTGATCGGCCCCACC-----CCGCTGAACATCATCGGCCCAACATGCTGACCCAGCTG 654  
Db 1684 TATTAGTAGGACCTTACCTACACCTGTCAACATAATTGGAAGAAATCTGTGACCCAGATC 1743  
Qy 655 GGCTGCACCTTGAACCTTCCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTCAAGGCC 714  
Db 1744 GGCTGCACCTTGAACCTTCCCCATCAGCCCTATTGAGACGGTGCCTGGAAGTTGAGCCG 1803  
Qy 715 GGCAATGACGCCCCCAAGGTGAAGCAGTGGGCCCTGACCGAGGAGAGATCAAGGCCCTG 774  
Db 1804 GGGATGACGCCCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAAGATCAAGGCCCTTA 1863  
Qy 775 ACCGCCATCTCGAGGAGATGGAGAAGAGGGGCAAGATCACCAGATCGGCCCGAGAAC 834  
Db 1864 GTCGAAATCTGTACAGAGATGGAGAAGGAGGGAAGATCAGCAAGATCGGCCCTTGAAAC 1923  
Qy 835 CCCTTACAACACCCCGCTGTTCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTG 894  
Db 1924 CCCTTACAACATCTCAGTCTTCGCCAATCAAGAAGAGGACAGTACCAAGTGGAGAAAGCTG 1983  
Qy 895 GTGACTTTCGCGAGCTGAAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954  
Db 1984 GTGGACTTCAGAGAGTGAACAAGAGAACTCAGGACTTCTGGGAAGTTTCACTGGGCATC 2043  
Qy 955 CCCACCCCGCGCTGAAGAAGAGAGAGAGCGTGCTGTGTGGAGCTGGGCGAGCGCC 1014  
Db 2044 CCACATCCCGTGGTGTGAAGAAGAAAGTCACTGACAGCTGTGTGGATGTGGGTGATGCC 2103  
Qy 1015 TACTTCAGCGTGCCTGGAGAGGACTTTCGCAAGTACACCGCTTCAACCATCCCGAGC 1074  
Db 2104 TACTTCTCCGTTCTTGGACGAGACTTTCAGGAAGTACACTGCTTCAAGATACCTTAGC 2163  
Qy 1075 ATCAACAACAGAGACCCCGCATCCGCTTACAGTACAAACGCTGTGCCCCAGGGCTGGAAG 1134  
Db 2164 ATCAACAACAGAGACACAGGATCCGCTACAGTACAAAGCTGTGCCACAGGATGAGAG 2223  
Qy 1135 GGAGCCCCAGACTTCTCAGAGCAGATGACCAAGATCTGTGGAGCCCTTCCGCGCCCGC 1194  
Db 2224 GGATCAGCAGCATCTTTCAAAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCAAGCAA 2283  
Qy 1195 AATCCCGAGATCGTGTATCAACA-----GGCCCCCTGTACGTGGGCGAGCGACTGGAG 1248  
Db 2284 AATCCAGACATCGTGTATCTATCAGTACATGAGCAGACTTACGTGAGGAAGTGACTGGAG 2343  
Qy 1249 ATC-GGGCAGCACCGCGCAAGATCGAGGAGCTCGCAAGCAACCTGTGTGGCTGGGGCTT 1307  
Db 2344 ATCGGGGACGACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGTGGGACT 2403  
Qy 1308 CACCAACCCCGAACAAGAGACCAAGAAGAGGCCCTTCTCTGTGTGATGGGCTACAGACT 1367  
Db 2404 GACCACACCAACAAGAAGACCAAGAGAACTCTCTTCTGTGTGATGGGCTACGAAC 2463  
Qy 1368 GCACCCGACAGTGGAGCGCTGAGGCCCATCGACTGCCGAGAGAGGAGCTGGACCGT 1427  
Db 2464 GCATCTGTACAGTGGGACAGTGGACCCCATCTGTCTGCTGAGAGAGGACAGCTGGACTGT 2523

Qy 1428 GAACGACATCCAGAAGCTGTGTGGCAAGCTGAACTGGGCGAGCCAGAGATCTTACCCCGGAT 1487  
Db 2524 GAACGACATCAGAAGCTGTGTGGCAAGTGTGAATGGGCGAAGCCAGATCTTACCCAGCAT 2583  
Qy 1488 CAAGTTCGCCAGCTGTGAAGCTGTCCGCGCGCAAGGCCCTGACCGACATCTGTGCC 1547  
Db 2584 CAAAGTTAGGCAGCTGTGCAAGCTGTTCGAGGAACCAAGGCACTGACAGAAGTGTATCCC 2643  
Qy 1548 CCTGACGAGGAGCCGAGCTGTGGAGCTGGCCGAGAACCGCGAGATCTCTGCCGAGCCCGT 1607  
Db 2644 ACTGACAGAGGAGCAGAGCTAGAACTTGCAGAGAACCGAGAGATCTCTGAAGGAGCCAGT 2703  
Qy 1608 GCACGGCTGTATCTACGACCCCAAGCAAGCACTGTGTGGCCGAGATCTCAGAAAGCAGGGCCA 1667  
Db 2704 ACATGGAGTGTACTTACGACCCCAAGCAAGCACTGTATCGCAGAGATCCAGAAAGCAGGGCA 2763  
Qy 1668 CGACCACTGACCTTACCAAGATCTTACCAAGGAGCCCTTCAAGAACCTTGAAGCCGCAAGTA 1727  
Db 2764 AGGCAATGGACCTTACCAAAATCTTACCAAGGAGCCCTTCAAGAACCTTGAAGCAGGCAAGTA 2823  
Qy 1728 CGCAGATGCGCACCCGCCCAACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAA 1787  
Db 2824 CGCAAGATGAGGGTGCCTCCACACCAACGATGTGAAGCAGCTGACAGGSCAGTGCAGAA 2883  
Qy 1788 GATCGCATGTGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTTGCCTATCCA 1847  
Db 2884 GATCACACAGAGAGCATCTGTGATCTGGGCAAGACTCCCAAGTTCAAGCTGCCCATACA 2943  
Qy 1848 GAGGAGACCTTGGGAGACCTTGTGGACCGACTACTGTGCGACCCACCTGGATTCGCCAGTG 1907  
Db 2944 GAAGGACATGAGGAGACATGTGTGACCCGAGTACTGGCAAGCCACTTGGATTCCTGTAGTG 3003  
Qy 1908 GGAATTCGTGAACACCCCTTGTGTGAAGCTGTGTGTACAGCTGGAAGAGGAGCCCAT 1967  
Db 3004 GGAATTCGTGAACACCCCTTGTGTGAAGCTGTGTGTATCAGCTGGAAGAGGAGCCAT 3063  
Qy 1968 CATCGCGCGCAGACCTTCTACGTGTGACGCGCGCCCAACCGCGAGACCAAGATCGGCAA 2027  
Db 3064 CGTGGAGCAGAGACCTTCTACGTGTGATGGGCGAGCCCAACAGGGAGACCAAGCTGGGCAA 3123  
Qy 2028 GCGCGCTAGCTGACGACCGGCGCGGAGAGATGTGTGAGCTGTGAGCTGTGAGACCCAGCA 2087  
Db 3124 GCGAGGCTAGCTGACCAACCGGAGACGACAGAAAGTGTGTGACCTGTGACACCAACCA 3183  
Qy 2088 CCAGAACCGCAGCTGACGACCTTCCAGCTGGCCCTGCAGGACAGCGGAGCAGCGAGGTGAA 2147  
Db 3184 CCAGAACCTGTAGCTGACGACCTTCTACCTAGCTTCTGCAAGACAGCGGACTGGAGGTGAA 3243  
Qy 2148 CATCGTACCGCAGCAGCTACGCTTGGGCTATCATCCAGGCCAGCCCGCAAGAGCGA 2207  
Db 3244 CATCGTACAGACTTCAAGTACG-CATGGGCTATCATCCAGACCAACAGACCAATCCGA 3302  
Qy 2208 GAGGAGCTGTGACACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTCTGAG 2267  
Db 3303 GTCAGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGGAGGTGTACTCTGGC 3362  
Qy 2268 CTGGGTCCCGCCCAAGGGCATCTGGCGGCAACGAGCAGATCGAACAGCTGTGTGAGCAA 2327  
Db 3363 ATGGGTACAGACACCAAGGAATTTGAGGAAATGAACAGTAGATTAATTAGTCAGTGC 3422  
Qy 2328 GGGCATCCGAAGGTGTGTCTCTGAGCGGATCGAT 2364  
Db 3423 TGGGATCCGGAAGGTGTCTTCTCTGGACCGGATCGAT 3459

RESULT 15  
US-09-952-060-1  
; Sequence 1, Application US/09952060  
; Patent No. 6733993  
; GENERAL INFORMATION:  
; APPLICANT: Emini, Emilio A.  
; APPLICANT: Youil, Rima

		; APPLICANT: Bett, Andrew J.		
		; APPLICANT: Chen, Ling		
		; APPLICANT: Kaslow, David C.		
		; APPLICANT: Shiver, John W.		
		; APPLICANT: Toner, Timothy J.		
		; APPLICANT: Casimiro, Danilo R.		
		; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS		
		; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND		
		; FILE REFERENCE: 20747Y		
		; CURRENT APPLICATION NUMBER: US/09/952,060		
		; CURRENT FILING DATE: 2001-09-14		
		; PRIOR APPLICATION NUMBER: PCT/US01/28861		
		; PRIOR FILING DATE: 2001-09-14		
		; PRIOR APPLICATION NUMBER: 60/317,814		
		; PRIOR FILING DATE: 2001-09-07		
		; PRIOR APPLICATION NUMBER: 60/279,056		
		; PRIOR FILING DATE: 2001-03-27		
		; PRIOR APPLICATION NUMBER: 60/233,180		
		; PRIOR FILING DATE: 2000-09-15		
		; NUMBER OF SEQ ID NOS: 38		
		; SOFTWARE: FastSeq for Windows Version 4.0		
		; SEQ ID NO 1		
		; LENGTH: 2577		
		; TYPE: DNA		
		; ORGANISM: Artificial Sequence		
		; FEATURE:		
		; OTHER INFORMATION: Codon optimized DNA encoding modified wt pol		
		US-09-952-060-1		
		Query Match 53.2%; Score 1309.6; DB 4; Length 2577;		
		Best Local Similarity 86.2%; Pred. No. 9.6e-215;		
		Matches 1463; Conservative 0; Mismatches 229; Indels 6; Gaps 1;		
QY	672	CCCCATCAGCCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCGATGAAGCGGCCCA	731	
DB	15	CCCCATCTCCCCATTGAAGCTGTGCCCTGTGAAGCTGAAGCTGGCATGGATGGCCCA	74	
QY	732	GGTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGSCCTGACCGCCATCTGCGAGGA	791	
DB	75	GGTGAAGCAGTGGCCCTGACTGAGAGAGAGATCAAGGCCCTGGTGGAAATCTGCACTGA	134	
QY	792	GATGGAGAGAGGGCAAGATCACAAGATCGGCCCGAGAACCCCTACAAACCCCGCT	851	
DB	135	GATGGAGAGAGGGCAAAATCTCAAGATTGGCCCGGAGAAACCCCTACAACACCCCTGT	194	
QY	852	GTTCGCCATCAAGAGAGAGAGACACCAAGTGGCGCAAGCTGTGTGAATCTCCGAGCT	911	
DB	195	GTTTGCCATCAAGAGAGAGAGATCCACCAAGTGGAGGAAGCTGTGTGAATCTCAAGGAGCT	254	
QY	912	GAAACAGCGCACCCAGGACTCTGGAGAGTGCAGCTGGGCATCCCCACCCCGCGGCCT	971	
DB	255	GAAACAGAGGACCCAGGACTCTGGAGAGTGCAGCTGGGCATCCCCACCCCGCTGGCCT	314	
QY	972	GAAAGAGAGAGAGCGTGACCGTCTGTGAAGTGGGCGACGCTTCTCAGCGTGGCCCT	1031	
DB	315	GAAAGAGAGAGAGTCTGTGACTGTCTGTGAAGTGGGAGTGCCTTCTCTGTGGCCCT	374	
QY	1032	GGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGAGACCC	1091	
DB	375	GGATGAGGACTTTCAGAGAGTACATGTCCTTCAACATCCCTCCATCAACAAATGAGACCCC	434	
QY	1092	CGGCATCCGCTACCAAGTCAACGCTGCGCCAGGGCTGGAAGGGCGACCCAGCATCTT	1151	
DB	435	TGGCATCAGGTACCAAGTACAAATGTCTGCCCGAGGGCTGGAAAGGGCTCCCTTGCATCTT	494	
QY	1152	CCAGAGCAGCATGACCAAGATCCTTGGAGCCCTTTCGCGCCCGCCAAACCCCGAGATCGTGAT	1211	
DB	495	CCAGTCTCTCATGACCAAGATCTCTGGAGCCCTTTCAGGAGAGAGAGCCCTTGACATTTGAT	554	
QY	1212	CTACCA-----GGCCCCCTGTACGTGGGAGAGACCTTGAGATTCGGCAGCACCGCC	1265	
DB	555	CTACCAAGTACATGATGACTGTATGTGGGCTCTGACCTGGAGATTTGGGAGAGACAGGAC	614	

Qy 2346 GTTCCTGGACGGCATCGA 2363  
|||||  
Db 1695 GTTCCTGGATGCGATTGA 1712

Search completed: June 2, 2005, 03:13:05  
Job time : 401 secs

**This Page Blank (uspto)**





Qy	398	GCC	CCCTG	TG	G	A	G	A	G	T	G	G	G	C	C	A	G	A	T	C	A	A	G	A	G	A	G	G	C	C	T	G	C	T	G	G	A	C	A	C	G	G	C	G	457
Db	241	G	C	C	C	C	T	G	T	G	A	C	C	T	C	A	G	A	T	C	A	G	A	G	A	G	G	G	C	T	G	C	T	G	C	A	C	A	C	G	G	C	G	300	
Qy	458	C	C	G	A	C	A	C	G	T	G	T	G	A	G	A	G	A	T	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	517		
Db	301	C	C	G	A	C	A	C	G	T	G	A	G	A	G	A	T	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	360		
Qy	518	G	C	G	C	A	T	C	G	G	G	T	T	C	A	T	C	A	A	G	T	G	C	G	C	A	G	A	C	A	G	A	T	C	C	T	G	A	T	C	G	G	577		
Db	361	G	C	G	G	A	T	C	G	G	G	T	T	C	A	T	C	A	A	G	T	G	C	G	C	A	G	A	C	A	G	A	T	C	C	C	G	T	G	A	T	C	G	420	
Qy	578	G	C	A	A	A	A	G	G	C	A	T	C	G	G	A	C	C	G	T	G	A	T	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	637		
Db	421	G	C	C	A	A	G	G	C	C	A	T	C	G	G	A	C	C	G	T	G	T	G	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	480		
Qy	638	A	C	A	T	G	T	G	A	C	C	C	A	G	T	G	C	A	C	C	T	G	A	A	T	T	C	C	C	C	A	T	C	A	G	C	C	C	C	A	G	C	697		
Db	481	A	C	T	G	T	G	A	C	C	A	G	T	C	G	T	G	C	A	C	C	T	G	A	A	T	T	C	C	C	C	A	T	C	A	G	C	C	C	A	G	A	G	540	
Qy	698	C	C	G	T	G	A	A	G	T	G	A	C	C	G	G	A	T	G	A	C	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	757		
Db	541	C	C	G	T	G	A	A	G	T	G	A	C	C	G	G	A	T	G	A	C	C	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	600		
Qy	758	A	G	A	A	T	C	A	A	G	C	C	T	G	A	C	C	C	A	T	C	T	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	G	A	C	A	817		
Db	601	A	G	A	A	T	C	A	A	G	C	C	T	G	T	G	A	T	C	T	G	A	C	C	G	A	T	T	G	G	A	A	G	A	G	A	G	A	G	A	G	A	660		
Qy	818	A	G	A	T	C	G	G	C	C	G	A	A	C	C	C	T	A	C	A	A	C	C	C	G	T	G	T	T	C	G	C	A	T	C	A	G	A	A	G	A	G	877		
Db	661	A	G	A	T	C	G	C	C	C	G	A	A	C	C	C	T	A	C	A	A	C	C	C	C	G	T	T	T	C	G	C	A	T	C	A	A	A	A	G	A	G	720		
Qy	878	C	C	A	A	T	C	G	C	G	A	C	T	G	T	G	A	C	T	T	C	C	G	A	G	T	G	A	A	C	A	A	C	G	C	C	C	A	G	A	C	T	937		
Db	721	C	C	A	A	T	C	G	C	G	A	C	T	G	T	G	A	C	T	T	C	C	G	A																					

Db	1321	GCATCAAGGTGAACGACGCTGTGAAGCTGTGGCGGCACCAAGGCCTCACCAGAGTGA	1380
Qy	1538	TGCCCTTGACCGAGAGGCGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTTGC CGCAGC	1597
Db	1381	TCCCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC	1440
Qy	1598	CCGTGCA CGCGTGTACTACGACCCAGCAAGGACTGTGTGGCGCGAGATCCAGAGCAGG	1657
Db	1441	CCGTGCA CGAGGTGTACTACGACCCAGCAAGGACTGTGTGGCGCGAGATCCAGAGCAGG	1500
Qy	1658	GCCACGACCACTGAGACCTTACCAGATCTACCAGAGGCCCTTCAAGAACCTCAAGACCGGCA	1717
Db	1501	GCCAGGCCAGTGGACCTTACAGATCTACCAGAGGCCCTTCAAGAACCTGAGACCGGCA	1560
Qy	1718	AGTACGCCAAGATGCGCACCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC	1777
Db	1561	AGTACGCCCGCATGCGCGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC	1620
Qy	1778	AGAAGATCGCATGGAGAGCATCGTGTACTTGGGGCAGACCCCAAGTTCGGCTGCACCA	1837
Db	1621	AGAAGTGGACCGCAGAGCATCGTGTACTTGGGGCAAGATCCCAAGTTCAAGCTGCCCA	1680
Qy	1838	TCCAGAAAGGAGACCTCGGAGACCTGTGTGACCGCACTACTGGCAGGCCACCTGGATCCCG	1897
Db	1681	TCCAGAGGAGACCTCGGAGGCCCTGTGTGATGAGTACTGCGAGGCCACCTGGATCCCG	1740
Qy	1898	AGTGGAGTTCTGTGAACACCCGCCCTGTGTGAAGCTGTGTACCAGCTGGAGAAAGAGC	1957
Db	1741	AGTGGAGTTCTGTGAACACCCGCCCTGTGTGAAGCTGTGTACCAGCTGGAGAAAGAGC	1800
Qy	1958	CCATCATCGCGCGCAGACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCG	2017
Db	1801	CCATCGTGGCGCGCAGACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGCTGG	1860
Qy	2018	GCAAGCCGGCTACTGTGACCGACCGGGCCGGCAGAAAGATCTGTAGCCTGACCGAGACCA	2077
Db	1861	GCAAGGCCGCTACTGTGACCGACCGGGCCGGCAGAAAGTGTGTAGCATCGCCGACACCA	1920
Qy	2078	CCAACCAAGAACCGAGCTGAGGCCATCCACTGGCCCTGAGGACAGCGGCTGAGG	2137
Db	1921	CCAACCAAGAACCGAGCTGAGGCCATCCACTGGCCCTGAGGACAGCGGCTGAGG	1980
Qy	2138	TGAACATCGTGAACGACAGCAGTACGCCCTGGGCATCATCCAGGGCCAGCCCGCAAGA	2197
Db	1981	TGAACATCGTGAACGACAGCAGTACGCCCTGGGCATCATCCAGGGCCAGCCCGCAAGA	2040
Qy	2198	GCGAGAGGAGCTGTGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAAAGTGTACC	2257
Db	2041	GCGAGAGGAGCTGTGTGAGCCAGATCATCGAGCAGCTGATCAAGAGGAGAAAGTGTACC	2100
Qy	2258	TGAGCTGGGTGCCCGCCCAACAAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGA	2317
Db	2101	TGGCCTGGGTGCCCGCCCAACAAGGGCATCGCGCGCAACGAGCAGTGTGAACAAGCTGTGA	2160
Qy	2318	GCAAGGCGATCCGCAAGGTGCTGTTCTGTGACGGCATCGATGCGCGCATCGTGTATCTACC	2377
Db	2161	GCGCGGCGATCCGCAAGGTGCTGTTCTGTGAA CGGCATCGATGCGCGCATCGTGTATCTACC	2220
Qy	2378	AGTACATGAGACCTGTAGCTGGGACGCGCGGCCCTAGGATCGATTTAAAGCTTCCCG	2437
Db	2221	AGTACATGAGACCTGTAGCTGGGACGCGCGGCCCTAGGATCGATTTAAAGCTTCCCG	2280
Qy	2438	GGGCTAGACCGGTGAATTC	2457
Db	2281	GGGCTAGACCGGTGAATTC	2300

RESULT 2  
US-09-475-515-82  
; Sequence 82, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475.515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 82  
; LENGTH: 2306  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(-).protmod.RTopt.YM  
US-09-475-515-82

Query Match 82.2%; Score 2019.2; DB 4; Length 2306;  
Best Local Similarity 93.2%; Pred. No. 0;  
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy	170	CGCGCAAGAGGGCCACAGATGAAGACTGCACCGAGCGCCAGGCCAATCTTCTCCGCG	229
Db	1	CGCGCGCGAGGACACCAATGAAGATTGCACTGAGAGACAGCTAAATTTCTCCGCG	60
Qy	230	AGGACCTGGCTTCCCGCAGGCGAGCCCGCGAGTTCCTCCAGGAGCAGAACCGCGCA	289
Db	61	AGGACCTGGCTTCTCGAGGCGAGCCCGCGAGTTGACGAGGAGCAGACCCGCGCA	120
Qy	290	ACAGCCCGCAGCGCGAGCTGAGGTGCGCGCGG-----ACAAACCCCGCAGCGAGG	343
Db	121	ACAGCCCGCAGCGCGAGCTGAGGTGCGCGCGGCGGAGAAACAAGCTGAGCGAGG	180
Qy	344	CGCGCGCGAGCGCGAGGCAACCTG-----AACTTCCCGCAGATCACTCTGGGAGC	397
Db	181	CGCGCGCGAGCGCGAGGCAACCTGAGCTTCAACTTCCCGCAGATCACTCTGGGAGC	240
Qy	398	GCCCCCTGGTACATCAAGTGGCGCGCAGATCAAGGAGGCGCTGCTGACACCGCG	457
Db	241	GCCCCCTGGTACATCAAGTGGCGCGCAGCTCAAGGAGGCGCTGCTGACACCGCG	300
Qy	458	CCGACGACACCGTCTGGAGAGATGAGCTTGGCGGCAAGTGAAGCCCAAGATGATCG	517
Db	301	CCGACGACACCGTCTGGAGAGATGAACCTTGGCGGCAAGTGAAGCCCAAGATGATCG	360
Qy	518	CGCGCATGGCGGTTTCATGAAGTGGCGCGCAGTACGACAGATCCTGATCGAGATCTCG	577
Db	361	CGCGCATGGCGGTTTCATGAAGTGGCGCGCAGTACGACAGATCCTCGTGGAGATCTCG	420
Qy	578	GCAAGAGGCCATCGGCACCGTGTGATCGCGGCCCAACCCCGTGAACATCATCTCGCGCA	637
Db	421	GCCAAGAGGCCATCGGCACCGTGTGATCGCGGCCCAACCCCGTGAACATCATCTCGCGCA	480
Qy	638	ACATGCTGACCCAGCTGGGCTGCACTTGAACCTTCCCGCATCAGCCCATCGAGACCGTGC	697
Db	481	ACCTGCTGACCCAGATCGGCTGCACTTGAACCTTCCCGCATCAGCCCATCGAGACCGTGC	540
Qy	698	CCGTGAAGCTGAAGCCCGGCTGAGCGGCCCAAGTGAAGCAGTGGCCCTTGAACCGAGG	757
Db	541	CCGTGAAGCTGAAGCCCGGCTGAGCGGCCCAAGTGAAGCAGTGGCCCTTGAACCGAGG	600
Qy	758	AGAAGATCAAGGCTTGAACCGCATCTGCGAGGAGATGAGAGAGGCGGCAAGATCACCA	817
Db	601	AGAAGATCAAGGCTTGTGGAGATCTGCAACCGAGATGAGAGAGGCGGCAAGATCACCA	660

Qy	818	AGATCGCCCGCGAAGAACCCCTTAAACACCCCGCTGTTCGCCATCAAGAAAGAGACACA	877
Db	661	AGATCGCCCGCGAAGAACCCCTTAAACACCCCGCTGTTCGCCATCAAGAAAGAGACACA	720
Qy	878	CCAAGTGGCGCAAGCTGGTGGACTTTCGGGAGCTGAAACAGCGCACCCAGACTTCTGGG	937
Db	721	CCAAGTGGCGCAAGCTGGTGGACTTTCGGGAGCTGAAACAGCGCACCCAGACTTCTGGG	780
Qy	938	AGGTGACAGCTGGGCGATCCCCACCCCGCGCGCTTGAAGAAAGAAAGAGCTGACCGTGC	997
Db	781	AGGTGACAGCTGGGCGATCCCCACCCCGCGCGCTTGAAGAAAGAAAGAGCTGACCGTGC	840
Qy	998	TGACAGTGGGCGACGCTTATTCAGCTGCGCTTGAAGAGGACTTTCGCAAGTACACCG	1057
Db	841	TGACAGTGGGCGACGCTTATTCAGCTGCGCTTGAAGAGGACTTTCGCAAGTACACCG	900
Qy	1058	CTTTCACATCCCCGAGCATCAACAGAGAGACCCCGCGCATCCGCTACAGTACACGTCG	1117
Db	901	CTTTCACATCCCCGAGCATCAACAGAGAGACCCCGCGCATCCGCTACAGTACACGTCG	960
Qy	1118	TGCCCCAGGGCTGAAGGGGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG	1177
Db	961	TGCCCCAGGGCTGAAGGGGAGCCCGCGCATCTTCCAGAGCAGCATGACCAAGATCTCTGG	1020
Qy	1178	AGCCCTTCCGCGCCGCAACCCCGAGATCTGTGATCTACAGGCGCCCTCTGTACGTCGCA	1237
Db	1021	AGCCCTTCCGCGCAAGAGAACCCCGAGCATCTGTGATCTACAGGCGCCCTCTGTACGTCGCA	1080
Qy	1238	GCGACCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGGAGCTTGGCAGCACCTGCTGC	1297
Db	1081	GCGACCTGGAGATCGGCGCAGCACCGCGCAAGATCGAGGAGCTTGGCAGCACCTGCTGC	1140
Qy	1298	GCTGGGCGCTTCAACACCCCGCAAGAGACCAAGAGGAGCGCCCTCTCTGCGCCAT--	1355
Db	1141	GCTGGGCGCTTCAACACCCCGCAAGAGACCAAGAGGAGCGCCCTCTCTGCGGATGG	1200
Qy	1356	-----CGAGCTGCAACCCCGCAAGTGAAGCTGTGAGCGCCATGAGCTGCGCGAGAGGAGA	1411
Db	1201	GCTACGAGCTGCAACCCCGCAAGTGAAGCTGTGAGCGCCATCATGCTGCGCGAGAGAGA	1260
Qy	1412	GCTGGACCGTGAACGACATCGAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCT	1471
Db	1261	GCTGGACCGTGAACGACATCGAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCT	1320
Qy	1472	ACCCCGGCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGGCGCGCAAGGCCCTGACCG	1531
Db	1321	ACCCCGGCATCAAGTGAAGCTGTGCAAGCTGTGCGGCGCGCAAGGCCCTGACCG	1380
Qy	1532	ACATCTGTCCTCGACCGAGGAGCGGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGC	1591
Db	1381	AGGTGATCCCTCGACCGAGGAGCGGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGA	1440
Qy	1592	CGAGACCGCTGACCGCGGTGTATCTACGACCGCAGCAAGAGCTGGTGGCGCAGATCTCGA	1651
Db	1441	AGGAGCCCGTGCACGAGGTGTATCTACGACCGCAGCAAGAGCTGGTGGCGCAGATCTCGA	1500
Qy	1652	AGCAGGCGCAGCAGCAGTGGACCTTACAGATCTACAGGAGCGCTTCAAGAACCTGGAAGA	1711
Db	1501	AGCAGGCGCAGCAGCAGTGGACCTTACAGATCTACAGGAGCGCTTCAAGAACCTGGAAGA	1560
Qy	1712	CCGCAAGTACCGCAAGATGCGCACCGCCCAACCAACAGAGCTGAAGCAGCTGACCGAGG	1771
Db	1561	CCGCAAGTACCGCCCGCATGCGCGCGCCCAACCAACAGAGCTGAAGCAGCTGACCGAGG	1620
Qy	1772	CCGTGCAAGAGATCGCCATCGAGAGCATCTGTGATCTGGGGCAAGACCCCGAAGTTCGCC	1831
Db	1621	CCGTGCAAGAGATCGCCATCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGTTCAGC	1680
Qy	1832	TGCCCATCCAGAGGAGACCTTGGGAGACCTGTGTGACCGGCTACTTGCAGGCGCACCTTGA	1891
Db	1681	TGCCCATCCAGAGGAGACCTTGGGAGGCGCTGTGTGATGAGTACTTGCAGGCGCACCTTGA	1740
Qy	1892	TCCCGGAGTGGAGTTCTGTGAACACACCCCGCCCTGTGTGAAGCTGTGTGATCCAGCTGGAGA	1951



QY 1232 TGGGAGCGGAGCTGGAGATCGGCGAGCAACCGCGCCAGAGTCGAGAGCTGGCGAGCACC 1291  
Db 1081 TGGGAGCGGAGCTGGAGATCGGCGAGCAACCGCGCCAGAGTCGAGAGCTGGCGAGCACC 1140  
QY 1292 TGCTGCGCTGGGGCTTACCAACCCCGGACAGAGAGCACCAGAGAGCCCGCTTCTGTC 1351  
Db 1141 TGCTGCGCTGGGGCTTACCAACCCCGGACAGAGAGCACCAGAGAGCCCGCTTCTGTC 1200  
QY 1352 CCAT-----CGAGCTGCACCCCGGACCAAGTGGACCGTGCAGCCCATCGAGTGGCCGAGA 1405  
Db 1201 GGATGGGCTACGAGCTGCACCCCGGACCAAGTGGACCGTGCAGCCCATCATGCTGCGCGAGA 1260  
QY 1406 AGGAGAGCTGACCGTGAACGACATCCGAGAGCTGGTGGGCAAGTGAATGGGCGAGCC 1465  
Db 1261 AGGAGAGCTGACCGTGAACGACATCCGAGAGCTGGTGGGCAAGTGAATGGGCGAGCC 1320  
QY 1466 AGATCTACCCCGGATCAAGGTGCGCAGCTGTCGAGCTGCTGGGGCGCCAGAGCC 1525  
Db 1321 AGATCTACCCCGGATCAAGGTGAAGCTGTGCAAGCTGTGCAAGCTGTGCGGGCACCAGGCGCC 1380  
QY 1526 TGACCGACATCGTGCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAGAACCGCGAGA 1585  
Db 1381 TGACCGAGGTGATCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAGAACCGCGAGA 1440  
QY 1586 TCCTGGCGGAGCGCTGACCGGCTGTACTAGACCCCGAGCAAGCACTGGTGGCGCGAGA 1645  
Db 1441 TCCTGAAGGAGCGCTGACCGGCTGTACTAGACCCCGAGCAAGCACTGGTGGCGCGAGA 1500  
QY 1646 TCCAGAGAGCGGCGCAGCAGCTGGACCTACGAGATCTACGAGAGCCCTTCAAGAAC 1705  
Db 1501 TCCAGAGAGCGGCGCAGGCGCAGTGGACCTTACGAGATCTTACGAGAGCCCTTCAAGAAC 1560  
QY 1706 TGAAGACCGGCAAGTACGCCAAGATGCGCCGCGCACCAACCAAGCTGAAGCAGCTGA 1765  
Db 1561 TGAAGACCGGCAAGTACGCCGCGATGCGCGCGCCACCAACCAAGCTGAAGCAGCTGA 1620  
QY 1766 CCGAGGCGCTGCAAGATGCCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGT 1825  
Db 1621 CCGAGGCGCTGCAAGATGAGCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCAAGT 1680  
QY 1826 TCCGCTGCCCATCCAGAGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCA 1885  
Db 1681 TCAAGCTGCCCATCCAGAGAGACCTGGGAGGCGCTGGTGGATGGAGTACTGGCAGGCCA 1740  
QY 1886 CCTGGATCCCGAGTGGGAGTTCTGTAAACACCCCGCTGGTGAAGCTGTGGTACCAGC 1945  
Db 1741 CCTGGATCCCGAGTGGGAGTTCTGTAAACACCCCGCTGGTGAAGCTGTGGTACCAGC 1800  
QY 1946 TGGAGAGGAGCCCATCATCGGCGCGAGACCTTTCTACGTGGACGGCGCGCCCAACCGCG 2005  
Db 1801 TGGAGAGGAGCCCATCGTGGGCGCGAGACCTTTCTACGTGGACGGCGCGCCCAACCGCG 1860  
QY 2006 AGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGCGAGAGATCGTAGCC 2065  
Db 1861 AGACCAAGCTGGGCAAGCGCGCTACGTGACCGACCGGGCGCGCGAGAGTGGTAGCA 1920  
QY 2066 TGACCGGAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGCGCTCGCAGGACA 2125  
Db 1921 TCGCCGACACCAACCAAGAGACCGAGCTGACGCGCATCCACTGGCGCGCGAGAGTGGTAGCA 1980  
QY 2126 GCGCAGCGAGGTGAACATCGTGAACGACAGCAGTACGCGCTGGGATCATCCAGGCGCC 2185  
Db 1981 GCGCGCTGGAGGTGAACATCGTGAACGACAGCAGTACGCGCTGGGATCATCCAGGCGCC 2040  
QY 2186 AGCCCGACAGAGCGAGAGCGAGCTGGTGAACGAGATCATGAGCAGCTGATCAAGAGG 2245  
Db 2041 AGCCCGACAGAGCGAGAGCGAGCTGGTGAACGAGATCATGAGCAGCTGATCAAGAGG 2100  
QY 2246 AGAAGTGTACTCAGCTGGTGGCGCCCGACAGGCGATCGGGCGGCAAGCAGAGATCG 2305  
Db 2101 AGAAGTGTACTCAGCTGGTGGCGCCCGACAGGCGATCGGGCGGCAAGCAGAGATCG 2160  
QY 2306 ACAAGCTGGTAGCAAGGCGATCCGCAAGGTGCTGTTCTTGGAGCGGATCGATGGCGGCA 2365

Db 2161 ACAAGCTGGTAGCGCGCGCATCCGCAAGGTGCTTCTTCAACGGCATCGATGGCGGCA 2220  
QY 2366 TCGTGATCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 2425  
Db 2221 TCGTGATCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCT 2280  
QY 2426 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2457  
Db 2281 AAAAGCTTCCCGGGCTAGCACCGGTGAATTC 2312

RESULT 4  
US-09-475-515-6  
; Sequence 6, Application US/09475515A  
; Patent No. 6602705  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: ZUR MEGEDE, Jan  
; APPLICANT: SRIVASTAVA, Indresh  
; APPLICANT: LIAN, Ying  
; APPLICANT: HARTOG, Karin  
; APPLICANT: LIU, Hong  
; APPLICANT: GREER, Catherine  
; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 4319  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: HIV-gag-polymerase  
US-09-475-515-6

Query Match 78.0%; Score 1915.8; DB 4; Length 4319;  
Best Local Similarity 89.4%; Pred. No. 0;  
Matches 2140; Conservative 0; Mismatches 227; Indels 28; Gaps 6;

QY 14 TGGCGAGGCGCATGAGCGAGCCAGCCAGC---GCCAACATCTCTGATGAGCGCAGCAACT 70  
Db 1100 TGGCGAGGCGCATGAGCGAGCCAGCCAGC---GCCAACATCTCTGATGAGCGCAGCAACT 1159  
QY 71 TCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGCGCCACATCGGCC 130  
Db 1160 TCGCAACAGCGGAGAGACCGTCAAGTGTCTTCAACTGCGGCAAGGAGGCGCCACACCGCA 1219  
QY 131 GCAACTGCG 190  
Db 1220 GGAAGTGCAGCG 1279  
QY 191 TGAAGGACTGCAGCG 250  
Db 1280 TGAAGGACTGCAGCG 1338  
QY 251 GCAAGGCG 310  
Db 1339 GGAAGGCG 1398  
QY 311 TGCAGGTGCG 364  
Db 1399 TGCAGGTGCG 1458  
QY 365 -----CCCTGAACTTCCCGCAGATCACTCTGTGGCAGCGCCCTCGTGAGCATCAAGG 418  
Db 1459 CTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACTAGGA 1518

Qy	419	TGGCGCGCAGATCAAGGAGCGCCTGCTTGACACCCGCGCCGACGACACCGTGTCTGGAGG	478
Db	1519	TCGCGCGCCAGCTCAAGAGGCGCTGCTGCACACCGCGCGCCGACGACACCGTGTCTGGAGG	1578
Qy	479	AGATGAGCCTGCCCGGCAAGTGGGAAGCCCAAGATGATCGGCGGCATCGGCGGCTTCATCA	538
Db	1579	AGATGAACCTGCCCGGCAAGTGGGAAGCCCAAGATGATCGGCGGCGATCGGGGGCTTCATCA	1638
Qy	539	AGGTGCGCCAGTACGACACAGATCCTGATTCGAGATCTCGCGCAAGAAAGGCCATCGGCACCG	598
Db	1639	AGGTGCGCGAGTACGACCAAGTCCCCTGGAGATCTGCGGCCCAAGGCCATCGGCACCG	1698
Qy	599	TGCTGATCGGCGCCACCCCCGTGAACATCATTCATCGGCGCGCAACATGCTGAACCCAGCTGGGCT	658
Db	1699	TGCTGTGTGGGCGCCACCCCGGTGAACATCATTCATCGGCGCGCAACCTGCTGACCCAGATGGCT	1758
Qy	659	GCACCTGAACTTCCCTCATCAGCCCCATCAGACCGTCAGACCGTCCCGTGAAAGTGAAGCCCGGCA	718
Db	1759	GCACCTGAACTTCCCTCATCAGCCCCATCAGACCGTTCAGACCGTTCGCGTGAAAGTGAAGCCCGGGA	1818
Qy	719	TGGAAGCCGCCAAGGTGAAGCAGTGGGCCCTTGACCGAGAGAAAGATCAAGGCCCTTGACCG	778
Db	1819	TGGAAGCCGCCAAGGTCAAGCAGTGGGCCCTTGACCGAGAGAAAGATCAAGGCCCTGGTGG	1878
Qy	779	CCATCTCGAGGAGATGGAGAAGAGGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCT	838
Db	1879	AGATCTGCACCGAGATGGAGAAGAGGGCAAGATCAGCAAGATCGGCCCCCGAGAACCCCT	1938
Qy	839	ACAACACCCCGCTGTCGCCATCAAGAAGAGAGACAGCACCAAGTGGCGCAAGCTGGTGG	898
Db	1939	ACAACACCCCGCTGTCGCCATCAAGAAGAGAGACAGCACCAAGTGGCGCAAGCTGGTGG	1998
Qy	899	ACTTCCCGAGCTGAACAAGCGCACCCAGAGCTTCTGGGAGGTGCAGCTGGGCATCCCC	958
Db	1999	ACTTCCCGAGCTGAACAAGCGCACCCAGAGCTTCTGGGAGGTGCAGCTGGGCATCCCC	2058
Qy	959	ACCCGCGGGCTGAAGAAGAGAAGAGCGTGACCGTGTGAGAGTGGGCGAGCGCTACT	1018
Db	2059	ACCCGCGGGCTGAAGAAGAGAAGAGCGTGACCGTGTGAGAGTGGGCGAGCGCTACT	2118
Qy	1019	TCAGCGTCCCTGAGAGAGACTTCCGCAAGTACACCGCTTCACCATCCCAGAGATCA	1078
Db	2119	TCAGCGTCCCTGAGAGAGACTTCCGCAAGTACACCGCTTCACCATCCCAGAGATCA	2178
Qy	1079	ACAAAGAGACCCCGGCATCCGCTACAGATACAAAGTGTGCTGCCCAAGGCTGGAAGGGCA	1138
Db	2179	ACAAAGAGACCCCGGCATCCGCTACAGTACAAAGTGTGCTGCCCAAGGCTGGAAGGGCA	2238
Qy	1139	GCCCGAGATCTTCCAGAGCAGATGACCAAGATCTTGAGAGCCCTTCGCGCCCGCAACC	1198
Db	2239	GCCCGGCACTTTCAGAGCAGATGACCAAGATCTCTGAGAGCCCTTCGCGCAAGCAACC	2298
Qy	1199	CCGAGATCTGTACTACCA-----GCCCCCTGTACGTGGGCGAGCGACTTGAGATCG	1252
Db	2299	CCGACATCTGTACTACAGTACATGAAGACCTGTAGTGGGCGAGCGACTTGAGATCG	2358
Qy	1253	GCCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCA	1312
Db	2359	GCCAGCACCGCACCAAGATCGAGAGCTGCGCCAGCACCTGCTGCGCTGGGGCTTCACCA	2418
Qy	1313	CCCCCGACAAGAGCACCAAGAGGAGCCCCCTTCCTGCCCAT-----CGAGCTGCACC	1366
Db	2419	CCCCCGACAAGAGCACCAAGAGGAGCCCCCTTCCTGCCCAT-----CGAGCTGCACC	2478
Qy	1367	CCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGCGAGAGAGAGCTGGACCGTGAACG	1426
Db	2479	CCGACAAGTGGACCGTGCAGCCCATCATGTGCCGCGAGAGAGACAGCTGGACCGTGAACG	2538
Qy	1427	ACATCCAGAGCTGGTGGGCAAGCTGAATCTGGGCGACGCCAGATCTACCCCGGATCAAGG	1486
Db	2539	ACATCCAGAGCTGGTGGGCAAGCTGAATCTGGGCGACGCCAGATCTACCCCGGATCAAGG	2598

Qy	1487	TGCGCCAGCTGTGGAAGACTCTCTCGCGCGCGCCCAAGGCCCTTGACCGACATCTGTTGCCCTCTGA	1546
Db	2599	TGAAGCAGCTGTGTCAAGACTCTCTCGCGGGCACCAAGAGCCCTTGACCGAGGTGATCCCCCTCTGA	2658
Qy	1547	CCGAGGAGGCCGAGCTGGAGCTGGCCGAGAAACCGCAGATCTCTCGCGAGACCCCGTGTGCACG	1606
Db	2659	CCGAGGAGGCCGAGCTGGAGCTGGCCGAGNAACCGCGAGATCCTTGAAGGAGCCCGTGTGCACG	2718
Qy	1607	GCGTGTACTACGACCCCGACAAGGAACCTGGTGTGCCGAGATTCAGAGACGAGGGCCACGACC	1666
Db	2719	AGGTGTACTACGACCCCGACAAGGACCTGGTGTGCCGAGATTCAGAAAGCAGGGCCGAGGGCC	2778
Qy	1667	AGTGGACCTACCAAGATCTACACGAGGCCCTTCAAGAACCTTGAAGACGGCAAGTACGCCCA	1726
Db	2779	AGTGGACCTACCAAGATCTACACGAGGCCCTTCAAGAACCTTGAAGACGGCAAGTACGCCCC	2838
Qy	1727	AGATGCGCACCGGCCACACCAACGACGCTGAAGCAGCTGACCGAGGCCGTGCAGAAGATCG	1786
Db	2839	GCATGCGCGCGCCACACCAACGACGCTGAAGCAGCTGACCGAGGCCGTGCAGAAGTGA	2898
Qy	1787	CCATGAGAGCATCTGTGATCTGGGGCAGAGACCCCCCAAGTTTCGCGCTGCCCATCCAGNAAGG	1846
Db	2899	GCACCGAGAGCATCTGTGATCTGGGGCAAGATCCCCAAGTTTCAAGCTGTGCCCATCCAGAAGG	2958
Qy	1847	AGACCTGGGAGACCTGGTGTGACCGCATCTATCTGCGAGCGCACCTCGATCCCCCGAGTGGGAGT	1906
Db	2959	AGACCTGGGAGGCGCTGGTGTGATGAGTACTGCGAGGCCACCTGGATCCCCCGAGTGGGAGT	3018
Qy	1907	TCGTGAACACCCCCCCTGGTGAAGCTGTGTGTGTAACAGCTGTGAAGAGGAGCCCATCATCTG	1966
Db	3019	TCGTGAACACCCCCCCTGGTGAAGCTGTGTGTGTAACAGCTGTGAAGAGGAGCCCATCTGTGG	3078
Qy	1967	GCGCCGAGACCTTCTAGTGTGACGGCGCGCCCAACCGCAGACCAAGATCGGCAAGGCCG	2026
Db	3079	GCGCCGAGACCTTCTAGTGTGACGGCGCGCCCAACCGCAGACCAAGCTGGGCAGAGGCCG	3138
Qy	2027	GCTACGTGACCGACCGGGCGCGCAGAGATTCGTGTAGCTTCGCGAGACCACCAACCGA	2086
Db	3139	GCTACGTGACCGACCGCGCGCGCCAGAAAGTGTGAGCATCGCCGACACCAACCGA	3198
Qy	2087	AGACCGAGCTGACGGCCATCCAGCTGGCCCTGCAGACACGGCAGCGAGGTGAAACATCG	2146
Db	3199	AGACCGAGCTGACGGCCATCCAGCTGGCCCTGCAGACACGGCCTGGAGGTGAAACATCG	3258
Qy	2147	TGACCGACACGCCAGTACGCCCTGGGCATCATTCAGGCCACCGCCGACAGGCGAGAGCG	2206
Db	3259	TGACCGACACGCCAGTACGCCCTGGGCATCATTCAGGCCACCGCCGACAGGCGAGAGCG	3318
Qy	2207	AGCTGTGAAACAGATCATCGACGAGCTGATCAAGAAGGAGAGGTGTACTGTAGCTGGG	2266
Db	3319	AGCTGTGTAGCCAGATCATCGACGAGCTGATCAAGAAGGAGAGGTGTACTGTGGCCTGGG	3378
Qy	2267	TGCCCGCCACAAAGGGCATCGCGGCCAAACGAGCAGATTCGACAAGCTGGTGTGAGCAAGGCCA	2326
Db	3379	TGCCCGCCCAAAAGGGCATCGCGGCCAAACGAGCAGGTGTGAGCAAGCTGGTGTGAGCGCCGCCA	3438
Qy	2327	TCGCAAGGTGCTGTTCTTCTGGAACGGCATTCGATGGCGGCATCTGTGTATCTACAGTGA	2381
Db	3439	TCGCAAGGTGCTGTTCTTCTTCTGAAACGGCATCCAGCAAGGCCCAAGGAGGAGCAGAGAA	3493

RESULT 5  
US-09-475-515-81  
: Sequence 81, Application US/09475515A  
: Patent No. 6603705  
: GENERAL INFORMATION:  
: APPLICANT: BARNETT, Susan  
: APPLICANT: ZUR WEGEDE, Jan  
: APPLICANT: SRIVASTAVA, Indresh  
: APPLICANT: LIAN, Ying  
: APPLICANT: HARTOG, Karin  
: APPLICANT: LIU, Hong  
: APPLICANT: GRBER, Catherine

; APPLICANT: SELBY, Mark  
; APPLICANT: WALKER, Christopher  
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
; FILE REFERENCE: 1621.002  
; CURRENT APPLICATION NUMBER: US/09/475,515A  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 81  
; LENGTH: 2299  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: FS(+) proinact.RTopt.YMMW  
US-09-475-515-81

Query Match 76.2%; Score 1872.8; DB 4; Length 2299;  
Best Local Similarity 89.3%; Pred. No. 2.2e-310;  
Matches 2055; Conservative 0; Mismatches 232; Indels 13; Gaps 3;  
QY 170 GCGCAAGGAGGCGCCACAGATGAAGGACTGCACCGAGCGCGAGCTTCCCGCG 229  
DB 1 CGGCGCGGAGGACACCAATGAAGATTGCATGAGAGAGAGCTTAATTTTT-AGGG 59  
QY 230 AGGACCTGGCTTCCCGAGGCGAAGGCCCGAGTTCCCGAGGAGCAGAACCGCGCA 289  
DB 60 AAGATCTGGCTTCTACAGGGAAGGCCAGGGAATTTTCTCAGAGCAGACGAGCA 119  
QY 290 ACAGCCACAGCGCGAGCTGCAGTGGCGGGG-----ACACCCCGCAGCGAGG 343  
DB 120 ACAGCCACAGAGAGAGCTTCAAGGTTTGGGAGGAGAAACAACTCCCTCTCAGAG 179  
QY 344 CCGCGCGGAGCGCCAGGCA-----CCCTGAATCCCGCAGATCACTCTGGCAGC 397  
DB 180 CAGGAGCGATAGACAGGAATGTATCTTTAACTTCCCTCAGATCACTCTTGGCAAC 239  
QY 398 GCCCCTCGTGAGCATCAAGTGGCGGCGCAGATCAAGGAGGCCCTGTGACACCGCG 457  
DB 240 GACCCCTCGTCAATAGATCGGGGGCACTCAGGAAGCGTGTCTCATACAGGAG 299  
QY 458 CCGACGACACCGTGTGAGAGATGAGCTTGGCGGCAAGTGAAGCCCAAGATGATCG 517  
DB 300 CAGATGATACAGTATTAGAGAAATGAATTTGGCAGGAAATGGAACCAAAATGATG 359  
QY 518 GCGCATCGCGGCTTCAATCAAGTGGCGCAGTACGACCATCTGATCAGATCTGCG 577  
DB 360 GGGGATCGGGGGCTTCAATCAAGTGGCGCAGTACGACCATCTGTAANAATCTGTG 419  
QY 578 GCAAGAGGCGCATCGGCACCGTGTGATCGGCGCCACCGCGTGAACATCATCGCGCGCA 637  
DB 420 GACATAAGCTATAGGTACAGTATTAGTAGACCTACACCTGTCAATATTGGAAGNA 479  
QY 638 ACATGCTGACCCAGTGGGTGCAACCTTGAATCTTCCCATCAGCCCATCAGACCGTGC 697  
DB 480 ATCTGTTGACCCAGATCGGTGCACTTGAATCTTCCCATCAGCCCATCAGACCGTGC 539  
QY 698 CCGTGAAGCTCAAGCCGCGATGAGCGCCCAAGTGAAGCTGAGCGCCCTGACCGAGG 757  
DB 540 CCGTGAAGTGAAGCCGCGATGAGCGCCCAAGTGAAGCTGAGCGCCATATGACCGAGG 599  
QY 758 AGAAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGAGAGAGGCGCAAGATCACCA 817  
DB 600 AGAAGATCAAGGCGCTTGTGAGATCTGCACCGAGATGAGAGAGGCGCAAGATCAGCA 659  
QY 818 AGATCGCGCCCGAGAAACCCCTTACACACCCCGGTGTTGCGCATCAAGAAGAGACAGCA 877  
DB 660 AGATCGCGCCCGAGAAACCCCTTACACACCCCGGTGTTGCGCATCAAGAAGAGACAGCA 719  
QY 878 CCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAAGCGCACCGGAGACTTCTGGG 937  
DB 720 CCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAAGCGCACCGGAGACTTCTGGG 779

QY 938 AGGTGCAGCTGGGCATCCCCACCCCGCGCCTGAAGAAGAAGAGCGTGAACGTCG 997  
DB 780 AGGTGCAGCTGGGCATCCCCACCCCGCGCCTGAAGAAGAAGAGCGTGAACGTCG 839  
QY 998 TGGAGTGGGCGAGCGCTTACTTACGCTGCGCTGGACGAGACTTCCGCAAGTACCG 1057  
DB 840 TGGAGTGGGCGAGCGCTTACTTACGCTGCGCTGGACGAGACTTCCGCAAGTACCG 899  
QY 1058 CTTTACCATCCCGAGCATCAACACGAGACCCCGGCATCCGCTTACAGTACACGTCG 1117  
DB 900 CTTTACCATCCCGAGCATCAACACGAGACCCCGGCATCCGCTTACAGTACACGTCG 959  
QY 1118 TGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGG 1177  
DB 960 TGCCCGAGGCTGGAAGGCGAGCCCGGCATCTTCCAGAGCAGCATGACCAAGATCTGG 1019  
QY 1178 AGCCCTTCCGCGCGCGACCCCGAGATCGTGATCTACAGGCGCCCTTACGTCGCGCA 1237  
DB 1020 AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACAGGCGCCCTTACGTCGCGCA 1079  
QY 1238 GCGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTCG 1297  
DB 1080 GCGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGAGCTGCGCGAGCACCTGTCG 1139  
QY 1298 GCTGGGCTTCCACACCCCGACAGAGCAACCAAGAGAGCGCCCTTCTCTGCCATCG 1357  
DB 1140 GCTGGGCTTCCACACCCCGACAGAGCAACCAAGAGAGCGCCCTTCTCTGCCATCG 1199  
QY 1358 AGCTGCACCCCGACAGTGGACCGTGACCGCATCGAGCTGCGCGAGAGAGAGCTGGA 1417  
DB 1200 AGCTGCACCCCGACAGTGGACCGTGACCGCATCTGCTGCGCGAGAGAGAGCTGGA 1259  
QY 1418 CCGTGAACGATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCG 1477  
DB 1260 CCGTGAACGATCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCG 1319  
QY 1478 GCATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCGCAAGCGCTGACCGACATCG 1537  
DB 1320 GCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCACCAAGCGCTGACCGAGTGA 1379  
QY 1538 TGCCCTTGAACGAGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGCGCGAGC 1597  
DB 1380 TCCCTTGAACGAGAGCGCGAGCTGGAGCTGGCGCGAGAACCGCGAGATCTCTGAGGAGC 1439  
QY 1598 CCGTGCAGCGGTGTACTACGACCCCGAGAGACTGTGTCGCGCGAGATCCAGAGCAGG 1657  
DB 1440 CCGTGCAGCGGTGTACTACGACCCCGAGAGAGCTGTGTCGCGCGAGATCCAGAGCAGG 1499  
QY 1658 GCCACGACGAGTGAACCTTACAGATCTACAGAGAGCGCTTCAAGAACCTGAAGACCGGCA 1717  
DB 1500 GCCAGGCGCAGTGAACCTTACAGATCTACAGAGAGCGCTTCAAGAACCTGAAGACCGGCA 1559  
QY 1718 AGTACGCGCAGATGCGCACCGCGCCACACCAACGAGTGAAGAGCTGACGAGCGCGTGC 1777  
DB 1560 AGTACGCGCAGATGCGCGCGCGCCACACCAACGAGTGAAGAGCTGACGAGCGCGTGC 1619  
QY 1778 AGAAGATGCGCATCGAGAGCATCTGTGATCTGGGCGAGAGCCCGCAAGTTCCGCTGCCCA 1837  
DB 1620 AGAAGTGAAGCACCGAGAGCATCTGTGATCTGGGCGAAGATCCCCAAGTTCAAGCTGCCCA 1679  
QY 1838 TCCAGAGAGGAGACTTGGGAGACTGTGTGAGACCGACTACTTGGCAGCGCACCTTGGATGCCCG 1897  
DB 1680 TCCAGAGAGGAGACTTGGGAGCGCTGTGTGATGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1739  
QY 1898 AGTGGAGTTCGTGAACACCCCGCGCGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1957  
DB 1740 AGTGGAGTTCGTGAACACCCCGCGCGCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1799  
QY 1958 CCATCATCGCGCGAGACTTCTACCTGAGCGCGCGCGCGCAACCGCGAGACCAAGATCG 2017  
DB 1800 CCATCGTGGCGCGAGACTTCTACCTGAGCGCGCGCGCGCGCAACCGCGAGACCAAGCTGG 1859





```
Qy 1356 ----CGAGCTGCACCCCGACAAGTGGACCGGTGCAGCCCATCGAGCTGCGGAGGAAGGAGA 1411
Db 1200 GCTACGAGCTGCACCCCGACAAGTGGACCGGTGCAGCCCATCGAGCTGCGGAGGAAGGAGA 1259
Qy 1412 GCTGGACCGGTGAACGAGCATCCAGAGCTGTTGGGCAAGCTGAATGGGCCAGCCAGATCT 1471
Db 1260 GCTGGACCGGTGAACGAGCATCCAGAGCTGTTGGGCAAGCTGAATGGGCCAGCCAGATCT 1319
Qy 1472 ACCCGCGCATCAAGGTGCGGCGAGCTGTGCAAGCTGTGCGGCGGCCAAGGCCCTGACCG 1531
Db 1320 ACGCGCGCATCAAGGTGAAGCATGTGCAAGCTGTGCGGCGGCCAAGGCCCTGACCG 1379
Qy 1532 ACATGTGTCCTTACCGAGGAGCCGAGCTGTGAGCTGTGCGGCGAAGCCGCGAGATCTGTC 1591
Db 1380 AGTGATCTCCCTGACCGAGGAGCCGAGCTGTGAGCTGTGCGGCGAAGCCGCGAGATCTGA 1439
Qy 1592 GCGAGCCGTGTACCGCGGTGTACTACGACCCAGCAGAGACCTGTGTCGCGGAGATCCAGA 1651
Db 1440 AGGAGCCGTGTACCGAGGTGTACTACGACCCAGCAGAGACCTGTGTCGCGGAGATCCAGA 1499
Qy 1652 AGCAGGCCAGCAGCAGTGGACCTTACGAGATCTACAGGAGCCCTTCAAGAACCTGAAGA 1711
Db 1500 AGCAGGCCAGCAGCAGTGGACCTTACGAGATCTACAGGAGCCCTTCAAGAACCTGAAGA 1559
Qy 1712 CCGCAAGTATCGCCCAAGATCGCACCGCCCAACACCAACGAGCTGAAGCAGCTGACCCGAGG 1771
Db 1560 CCGCAAGTATCGCCCGCATGCGCGGCCCAACCAACGAGCTGAAGCAGCTGACCCGAGG 1619
Qy 1772 CCGTGCAGAGATCGCCCATGTGAGAGCATCTGTGATCTGCGGCGAAGCCCAAGTTCGCC 1831
Db 1620 CCGTGCAGAGAGGTGAGCACCGAGAGCATCTGTGATCTGCGGCGAAGTCCCAAGTTCAGC 1679
Qy 1832 TGCCCATCCAGAGGAGACCTGGGAGACCTGTGTGACCGGACTACTGCGAGGCCACCTGGA 1891
Db 1680 TGCCCATCCAGAGGAGACCTGGGAGCCCTGTGTGATGGAGTACTTGGAGGCCACCTGGA 1739
Qy 1892 TCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTACAGCTGGAGA 1951
Db 1740 TCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTACAGCTGGAGA 1799
Qy 1952 AGAGCCCATCATCGGCGCCGAGACCTTTCTAGTGGACGGGCGCCGCAACCGCGAGACCA 2011
Db 1800 AGAGCCCATCATGTGGGCGCCGAGACCTTTCTAGTGGACGGGCGCCGCAACCGCGAGACCA 1859
Qy 2012 AGATCGGCAAGCGCGCTACGTGACCGACCGGGGCGGCGAGAGATCGTGAGCTGACCG 2071
Db 1860 AGTGGCAAGCGCGCTACGTGACCGACCGGGGCGGCGAGAGAGGTGTGAGCATCGCCG 1919
Qy 2072 AGACCAACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGACAGCGGCA 2131
Db 1920 ACACCAACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGACAGCGGCC 1979
Qy 2132 GCGAGGTGAACATCGTGACCGACAGCAGTACGCGCTTGGGATCATCCAGGCCAGCCCG 2191
Db 1980 TGGAGGTGAACATCGTGACCGACAGCAGTACGCGCTTGGGATCATCCAGGCCAGCCCG 2039
Qy 2192 ACAGAGCGAGCGAGCTGTGNAACGAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2251
Db 2040 ACAAGAGCGAGCGAGCTGTGAGCGAGATCATCGAGCAGCTGATCAAGAGGAGAGG 2099
Qy 2252 TGTACCTGAGCTGGGTCCCGCCCAAGGCGATCGGCGGCAACGAGCAGATCGACAAGC 2311
Db 2100 TGTACCTGGCTGGGTCCCGCCCAAGGCGATCGGCGGCAACGAGCAGATCGACAAGC 2159
Qy 2312 TGTGAGCAAGGGCATCCGCAAGTGTCTTCTGGAAGCGATCGATGGCGGATCGTGA 2371
Db 2160 TGTGAGCGCGGATCCGCAAGTGTCTTCTGGAAGCGATCGATGGCGGATCGTGA 2219
Qy 2372 TCTACCAAGTACATGGAAGCATCTGTACGTGGGAGCGGCGGCGGATCGATCGATTAAAGC 2431
Db 2220 TCTACCAAGTACATGGAAGCATCTGTACGTGGGAGCGGCGGCGGATCGATTAAAGC 2279
```

```
Qy 2432 TTCCCGGGGTAGCACCGGTGAATTC 2457
Db 2280 TTCCCGGGGTAGCACCGGTGAATTC 2305

RESULT 7
US-09-552-950-2
; Sequence 2, Application US/09552950
; Patent No. 6541248
; GENERAL INFORMATION:
; APPLICANT: Oxford Biomedica (UK) Limited
; TITLE OF INVENTION: Anti-Viral Vectors
; FILE REFERENCE: 674524-2004
; CURRENT APPLICATION NUMBER: US/09/552.950
; CURRENT FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4307
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:gagpol-synnp - codon
; OTHER INFORMATION: optimised gagpol sequence
US-09-552-950-2

Query Match 66.2%; Score 1626.6; DB 4; Length 4307;
Best Local Similarity 82.2%; Pred. No. 2.le-268;
Matches 1951; Conservative 0; Mismatches 394; Indels 28; Gaps 6;

Qy 12 CATGGCGGAGCCATGAGCCAGG---CCACGAGCGCCAAACATCTGTATGTCAGCGCAGCAA 68
Db 1086 CTTGGCTGAGGCCATGAGCCAGGTCACCACTCCGCTACCATCATGATGTCAGCGCGCAA 1145
Qy 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGCGGCAAGAGGAGGCCACATCGC 128
Db 1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGCTTCAACTGTCGCAAGAGAGGCCACAC 1205
Qy 129 CCGCAACTGCGGCGCCCGCCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGAGGCCACCA 188
Db 1206 CCGCAACTGAGGCGCCCTAGGAAAAGGCTGCTGGAATTCGCGCAAGAGAGGCCACCA 1265
Qy 189 GATGAAGGACTGACCCGAGCGCCAGGCAACTTCTTCCGCGAGGACCTTGGCTTCCCCCA 248
Db 1266 GATGAAGGACTGTACGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCTTCTTACA 1324
Qy 249 GGGCAGGCGCCGAGTTCCTCCAGCGAGAGAACCGCGCCCAACAGCCCGCCACCGCGGA 308
Db 1325 AGGGAAGGCCAGGGAATTTTCTTTCAGAGCAGACCCGCGCCCAACAGCCCGCGCA 1384
Qy 309 GCTGCAGGT-----GGCGCGGACCAACCCCGCAGCGAGGCGCGCGCGCCAGG 362
Db 1385 GCTTCAGGTCTGGGGTGCAGCAACAACTCCCCCTCCGAAGCAGGAGCCGCGCCAGG 1444
Qy 363 CA-----CCCTGAACCTTCCCCCAGATCACCTGTGCGAGCGCCCTTGTGAGCATCAA 416
Db 1445 CACGCTGTCTTCAACTTCTTCCCTCAGGTCAAGCTTTCGCGAGACCCCTCTGTACCATCA 1504
Qy 417 GGTGGCGCGCAGATCAAGAGAGCCCTGTGTGACACCGCGCGCGCAACAGACCTGTGTGA 476
Db 1505 GATCGGGGGGAGCTCAAGAGAGCTCTCTCTGACACCGCGAGCAGACGACCGTGTGGA 1564
Qy 477 GGAGATGAGCTCCCGCGCAAGTGGAGCCCAAGATGATCGGGGCAATCGCGGCTTCAT 536
Db 1565 GGAGATGTCTGTTCAGGCGCTGGAAGCCGGAAGATGATCGGGGGAATCGCGGCTTCAT 1624
Qy 537 CAAGGTGCGCCAGTACACCAAGATCTGTGATCGAGATCTGGCGCAAGAGGCAATCGGCGAC 596
Db 1625 CAAGGTGCGCCAGTATGACCAAGATCTCATCGAAATCTGGCGCAACAGGCTATCGGTAC 1684
Qy 597 CGTGTGATCGGCGCCACCCCGTGGAACATCATCGGCGCGCAACATGTGTGACCGAGTGGG 656
Db 1685 CGTGTGTGGGGCCCCACACACCCCGTCAACATCATCGGAGCGCAACCTGTGTGACGAGATCGG 1744
```



Matches 1939; Conservative 0; Mismatches 406; Indels 28; Gaps 6;

```
Qy 12 CATGCCGAGGCCATGAGCAGG---CCACAGCGCCAAACATCTGATGACGCGCAGCAA 68
Db 1086 CTTGGCTGAGGCCATGAGCGAGGTGACCAATCCGCTACCATCATGATGACGCGCGCAA 1145
Qy 69 CTTTCAGGGCCCCAGCGCATCATCAAGTGTCTCAACTGGCGCAAGAGGGCCACATCGC 128
Db 1146 CTTTCGGAACCAACGCAAGATCGTCAAGTGTCTCAACTGTGGCAAGAGGGCACACAGC 1205
Qy 129 CCGCAACTGCGCGCCCCCGCAAGAGGGCTGTGGAAGTGGCGCAAGAGGGSCACCA 188
Db 1206 CCGCACTGCAAGGCCCTTAGGAAGAGGGCTGTGGAATGCGCGAAGAGGCCACCA 1265
Qy 189 GATGAAGACTGCAACCGAGCGCCAGGCGCAACTCTTCCGCGAGGACCTGGCCTTCCGCCA 248
Db 1266 GATGAAGACTGTACTGAGAGACAGGCTAA-TTTTATTAGGGAAGATCTGGCCTTCTTACA 1324
Qy 249 GGGCAAGCGCGGAGTTCCCAAGAGCAGAACCGCGGCCCAAGAGGCCCGCGCA 308
Db 1325 AGGGAAGGCCAGGGAAATTTCTTCAGAGCAGACACAGAGCCCAACAGAGAGAGA 1384
Qy 309 GCTGCAGGTGCGCGG-----CGACAACCCCGCAGCGAGGCGCGCGCGAGGCCAGGG 362
Db 1385 GCTTCAGGTCTGGGGTAGAGAACAACTCCCTCAGAAAGCAGAGCCGATGACAAGG 1444
Qy 363 CA-----CCCTGAATCTCCCAAGATCACCTGTGGCAGCGCCCTCTGTGAGCATCAA 416
Db 1445 AACTGTATCTTTAACTTCCTCAGATCACTCTTTGGCAAGGACCTCTGTCAATATA 1504
Qy 417 GTTGGGCGCGCAGATCAAGAGAGGCCCTGTGTGAACAACCGCGCGCGCAACAGACCGTGTGA 476
Db 1505 GATAGGGGGCGAGCTCAAGAGAGGCTCTCTGTGAACAACCGAGCAGACACCGTGTGGA 1564
Qy 477 GGAGATGAGCTCGCGCGCAAGTGGAGCCCAAGATGATCGGCGCATCGCGCGCTTCAT 536
Db 1565 GGAGATGCTGTGTGCGAGCGCGTGGAAAGCCGAAGATGATCGGGGGAATCGGCGGTTTCAT 1624
Qy 537 CAAGGTGCGCGCAGTACACACAGATCTCTGATCGAGATCTGGGGCAAGAGGCCATCGGCAC 596
Db 1625 CAAGGTGCGCGCATGACACAGATCTCTCATCGAATCTGCGGCCAAGAGGCTATCGGTAC 1684
Qy 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGG 656
Db 1685 CGTGTGTGGCGGCCACACCCCGTCAACATCATCGAGCGCAACCTGTTGACGAGATCGG 1744
Qy 657 CTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAGCGCGG 716
Db 1745 TTGCACGCTGAACCTTCCCATTAGCCCTATCGAGACCGTACCGGTGAAGCTGAAGCCCGG 1804
Qy 717 CATGAGCGGCCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGAGAAAGATCAAGGCCCTGAC 776
Db 1805 GATGGAACGCGCGAAGGTCAAGCATATGCGCATTTGACAGAGGAGAAAGATCAAGGCACTGGT 1864
Qy 777 CGCATCTGCGAGGATGAGAAAGAGGGCAAGATCACCAAGATCGGCCCGCGAGAACCC 836
Db 1865 GGAGATTTGCACAGATGGAAGAGGAAGGAAATCTCCAAGATTGGGCTCGAGAACCC 1924
Qy 837 CTACAAACCCCGGTGTTCCCATCAAGAGAGGACAGACCAAGTGGCGCAAGCTGGT 896
Db 1925 GTACAAACCGCGGTGTTCCCAATCAAGAGAGGAGACTCGACGAAATGGGCGCAAGCTGGT 1984
Qy 897 GGACTTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGCATCCC 956
Db 1985 GGACTTCGCGAGCTGAACAGCGCACCGAAGACTTCTGGGAGGTTCAGCTGGGCATCCC 2044
Qy 957 CCACCCCGCGGCTGAAGAGAAAGAGCGTGAACCGTGTGGAACGTGGGCGAGCGCTA 1016
Db 2045 GCACCCCGCAGGGCTGAAGAGAAAGAAATCCGTGACCGTACTGTGATGTGGGTGATGCTA 2104
Qy 1017 CTTCAAGCGTGCCTGACGAGGACTTCCGCAAGTACACCGCCTTCAACATCCCCAGCAT 1076
Db 2105 CTTTCCTCGGTTCCCTGACGAAAGACTTCAGGAAGTACACTGCGCTTCAATTCCTTCGAT 2164
```

```
Qy 1077 CAACAACGAGACCCCGGSCATCGCTACAGTACCAAGTGTCTGCCCCAGGGCTGGAAGGG 1136
Db 2165 CAACAACGAGACACCGGGGATTCGATATCAGTACAACTGCTGCCCCAGGGCTGGAAGG 2224
Qy 1137 CAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGCGCTTCGCGGCCCGCAA 1196
Db 2225 CTCTCCCGCAATCTTCAGAGTAGCATGACCAAAATCTCGAGCCTTTCCGCAACAGAA 2284
Qy 1197 CCGCGAGATCGTGTATCTACCA-----GGCCCCCTGTAGTGGGCGAGCGACCTGGAGAT 1250
Db 2285 CCGCGACATCGTCATCTATCAGTACATGATGACTTGTACGTGGGCTCTGATCTAGAGAT 2344
Qy 1251 CGGCGACGACCGCGCCAGATCGAGGAGCTGCGCAAGCACTGTGTGCGTGGGCTTCAC 1310
Db 2345 AGGCGACGACCGCAACCAAGATCGAGGAGCTGCGCAGCACCTGTTGAGGTGGGACTGAC 2404
Qy 1311 CACCCCCGCAAGAGACCAAGAGGAGCCCCCTCTCTGCCCCAT-----CGAGCTGCA 1364
Db 2405 CACACCCGCAAGAGACCAAGAGGAGCTCCCTCTCTGGATGGGTTACGAGCTGCA 2464
Qy 1365 CCGCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAAAGGAGAGCTGGAACGCGAA 1424
Db 2465 CCTGACAAATGGAACCGTGCAGCTATCTGTCTGCGCAGAGAAAGACAGCTGGAATGTC 2524
Qy 1425 CGACATCCAGAAGCTGTGGGCAAGCTGAACTGGGCGCAGCCAGATCTACCCCGGCATCAA 1484
Db 2525 CGACATACAGAAGCTGTGGTGGGAAGTTGAACTGGGCGCAGTCAGATTTACCCAGGGATTAA 2584
Qy 1485 GGTGCGCAGCTGTGCAAGCTGTGCGCGGGCGCAAGCCCTGACCGACATCTGTCGCCCT 1544
Db 2585 GGTGAGGAGCTGTGCAAACTCTCTCGCGGAACAAGGCACTCAAGAGGTGATCCCCCT 2644
Qy 1545 GACCGAGGAGCGCGAGCTGAGCTGGCGGAGAACCGGAGATCTCTGCGAGCCGCTGCA 1604
Db 2645 AACCGAGGAGCGCGAGCTCGAACTGGCAGAAACCGAGAGATCTCTAAAGAGCCCGTGA 2704
Qy 1605 CGGCGTGTACTAGACCCCGCAGCAAGGACCTGTGGTGGCGGAGATCCAGAGCAGGCGCAG 1664
Db 2705 CGGCGTGTACTATGACCCCTCCAAAGGACCTGATCGCGGAGATCCAGAGCGAGGCGCAAG 2764
Qy 1665 CAGTGAACCTACAGATCTACAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTAGCG 1724
Db 2765 CAGTGAACCTATCAGATTTTACAGGAGCCCTTCAAGAACCTGGAAGACCGGCAAGTAGCG 2824
Qy 1725 CAAGTGCACCGCGCCACACACAGCAGCTGAAGCAGCTGACCGAGCGCGTGCAGAAAGAT 1784
Db 2825 CCGATGAGGGGTGCCCACTAACGACGCTAACGACGCTGACCGAGCGCGTGCAGAAAGAT 2884
Qy 1785 CGCATGGAGAGCATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAA 1844
Db 2885 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCTCTAAGTTCAAGTGCCTCATCCAGAA 2944
Qy 1845 GGAAGCTGGGAGACCTGTGTGAGCAGCTACTTGGCAGGCGCACTGGATCCCGGATGGGA 1904
Db 2945 GGAAGCTGGGAAACCTGTGTGAGCAGATATGCGAGGCCCACTGGGATTCCTGAGTGGGA 3004
Qy 1905 GTTGTGAAACACCCCGCTGTGTGAAAGCTGTGTACAGCTGGAAGAGGAGGCCATCAT 1964
Db 3005 GTTGTGTCAACCCCTCTCCCTGTGTGAAGCTGTGTATCAGCTGGAGAGGAGGCCATAGT 3064
Qy 1965 CGGCGCGGAGACCTTCTTACGTGGAACCGCGCGCCCAACCGCGAGAGCAAGATCGGCAAGGC 2024
Db 3065 GGGCGCGGAAACCTTCTACGTGGAAGGCGCGCTAAACAGGAGACTAAGCTGGGCAAGGC 3124
Qy 2025 CGGCTAGCTGACGACCGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAACCA 2084
Db 3125 CGGATACGTCACTAAACCGGGGCGAGACAGAAAGTTGTCACTCTCACTGACACCAACCA 3184
Qy 2085 GAAGACCGAGCTGACGAGCCATCCAGCTGGCCCTCGAGGAGACAGCGGAGGTTGAACAT 2144
Db 3185 GAAGACTGAGCTGACGAGCCATTTACCTCGCTTTTGCAGGACTCGGGGCTTGGAGGTGAACAT 3244
```



Db 2545 CGACATACAGAGCTGTGGGAAAGTTGAACCTGGGCCAGTCCAGATTTATCCAGGGATTAA 2604  
Qy 1485 GGTGGCCAGCTGTCAAGCTGTCTGCGGGGGCCAAAGGCCCTGACCGACATCTGCGCCCT 1544  
Db 2605 GGTGAGGAGCTGTGCAAACTCTCTCGCGGAACCAAGGCACTCAAGAGGTGATCCCT 2664  
Qy 1545 GACGAGAGAGCGAGCTGTGAGCTGTGGCGGAGAAACCGCGAGATCTCTGCGGAGCGCGTGCA 1604  
Db 2665 AACCGAGAGGCGGAGCTGTGAGCTGTGGCAGAAACCGAGAGATCTCTAAAGGAGCGCGTGCA 2724  
Qy 1605 CGGCTGTGTACTACGACCCACCAAGGACCTGTGGCGGAGATCCAGAGCAGAGGCGCCAGA 1664  
Db 2725 CGGCTGTGTACTACGACCCCTCAAGAGCTGTGCGCGAGATCCAGAGGAGGCGCAAG 2784  
Qy 1665 CCAGTGGACCTACAGATCTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGC 1724  
Db 2785 CCAGTGGACCTATCAGATTTACAGGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGC 2844  
Qy 1725 CAAGATGGACCGCCACCAAGAGCTGTGAAGAGCTGACCGAGGCGGTGCGAGAGAT 1784  
Db 2845 CCGATGAGGGGTGCCCACTAAACGAGCTCAAGCAGCTGACCGAGGCGGTGCGAGAGAT 2904  
Qy 1785 CGCCATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAA 1844  
Db 2905 CACACCGAAGCATCTGTGATCTGGGAAAGACTCTCAAGTTCAAGCTGCCCCATCCAGAA 2964  
Qy 1845 GGAGACCTGGGAGACCTGTGTGAGACGACTGTGGCAGGCACTTGGATCCCGGAGTGGGA 1904  
Db 2965 GGAACTCTGGAAACCTGTGTGAGCAGAGATTTGGCAGGCACTTGGATCTCTGAGTGGGA 3024  
Qy 1905 GTTCGTGAACACCCCTGCTGTGAAGCTGTGTGTAACAGCTGAGAGAGGAGCCCATCAT 1964  
Db 3025 GTTCGTGAACACCCCTGCTGTGTGAAGCTGTGTGTAACAGCTGAGAGAGGAGCCCATAGT 3084  
Qy 1965 CGGCGCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2024  
Db 3085 GGGCGCGAAGACCTTCTACGTGGATGGGCGCGCTAAACAGGGAGACTAAGCTGGGCAAGC 3144  
Qy 2025 CGGCTAGCTGACCGAGCCGGGCGGAGAGATCGTGAAGCTGTGACCGAGACCAACCA 2084  
Db 3145 CGGATAGCTACTAAACCGGGGAGACAGAGAGTTGTACCTCTACTGACACCAACCA 3204  
Qy 2085 GAAGACGAGCTGACGGCATCTCAAGCTGGCCCTCGAGGACAGCGGACGAGGTGAACAT 2144  
Db 3205 GAAGACTGAGCTGACGGCATTTACCTCGCTTTGACGAGACTCGGCGCTGGAGGTGAACAT 3264  
Qy 2145 CGTGACGACAGCAGTACGCGCTGGGATCATCCAGGCGGCGGCGGCGGCGGCGGCGGAG 2204  
Db 3265 CGTGACAGACTCTCAGTATGCGCTGGGATCATTTCAAGCGGCGGCGGCGGCGGCGGAGTGC 3324  
Qy 2205 CGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGCTG 2264  
Db 3325 CGAGCTGTGATCATGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTATCTTGGCGCTG 3384  
Qy 2265 GGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2324  
Db 3385 GGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3444  
Qy 2325 CATCGGCAAGGTGTCTTCTGGAGCGGATCGA 2357  
Db 3445 CATCAGGAAGGTGTCTTCTGGATGCGATCGA 3477

## RESULT 10

US-09-936-572-13

; Sequence 13, Application US/09936572

; Patent No. 6783981

; GENERAL INFORMATION:

; APPLICANT: UDEN, MARK

; APPLICANT: MITROPHANOUS, KYRIACOS

; TITLE OF INVENTION: ANTI-VIRAL VECTORS

; FILE REFERENCE: 078883/0137

; CURRENT APPLICATION NUMBER: US/09/936,572

; CURRENT FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: PCT/GB00/01002  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: GB 9906177.2  
; PRIOR FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4353

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURES:

; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP3-codon

; OTHER INFORMATION: Optimised HIV-1 gagpol with leader sequence from

; OTHER INFORMATION: the major splice donor

US-09-936-572-13

Query Match 65.1%; Score 1599.4; DB 4; Length 4353;

Best Local Similarity 81.5%; Pred. No. 9e-264;

Matches 1934; Conservative 0; Mismatches 411; Indels 28; Gaps 6;

Qy 12 CATGCGCGAGGCCATGAGCCAGG---CCACCAGCGCCAAACATCTCTGATGCGAGCGCAGCAA 68  
Db 1132 CTGCGTGAAGCCATGAGCCAGGTGACCAACTCCGCTACCATCATGATGCGAGCGCGCAA 1191  
Qy 69 CTTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGGAGGCGCCACATCGC 128  
Db 1192 CTTTCGGAACCAACGCAAGATCGTCAAGTGTCTCAACTGTTGCAAGAGAGGCGCACAGC 1251  
Qy 129 CGCAACTGCGCGCCCGCCCGCAAGAGGGCTGTGAAAGTGTGCGCAAGAGGCGCACCA 188  
Db 1252 CCGCAACTGCGAGGCGCCCTAGGAAAAGGGCTGTGAAAATGTGAAAAGAGGACACCA 1311  
Qy 189 GATGAAGAGCTGCAACCGAGCGCCAGGCAACTCTTCGCGAGGACCTGGGCTTCCCGCA 248  
Db 1312 AATGAAAGATTGTACTGAGAGACAGGCTAA---TTTTTAGGAAGATCTTGGCTTCCCGCA 1370  
Qy 249 GGGCAAGCGCCCGGAGTTCGCCAGGAGCAGAACCGCGCCAAACAGCCCGCCACAGCGCGCA 308  
Db 1371 AGGGAAGCGCCAGGAAATTTCTTTCAGAGCAGACAGGAGCCAGACCGCCCGCCAGAGAGA 1430  
Qy 309 GCTGAGGTGCGCGG-----CGACAAACCCCGCAGCGAGGCGCGGCGGCGGCGGCGG 362  
Db 1431 GCTTCAGGTGTGGGAAAGAGACAACTCCCTCTCAGAAAGCAGGAGCGGATAGACAAG 1490  
Qy 363 CA-----CCCTGAATTTCCCGCAGATCACCTGTGCGAGCGCCCGCTGTGAGCATCAA 416  
Db 1491 AACTGTATCTTTAGCTTCCCTCAGATCACTCTTTTGGCAGCGACCCCTCGTCACAATAA 1550  
Qy 417 GGTGGGCGGCGCAGATCAAGAGGCGCTGTGGACACCGGCGCGCAGACACACCGGTGCTGGA 476  
Db 1551 GATAGGGGGGCGAGCTCAAGGAGGCTCTCTTGGACACCGGAGCAGACGACCGGTGCTGGA 1610  
Qy 477 GGAGATGAGCCTGCGCGCAAGTGGAAAGCCCAAGATGATCGGCGGCAATCGGCGGCTTCAT 536  
Db 1611 GGAGATGCTGTGCGAGCGCGCTGGAAGCCGAGATGATCGGCGGAAATCGGCGGCTTTCAT 1670  
Qy 537 CAAGGTGCGCAGTACGACAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATCGGCAC 596  
Db 1671 CAAGGTGCGCAGTATGACAGATCTCTCATCGAAATCTGCGGCGCACAAGGCTTATCGGTAC 1730  
Qy 597 CGTGTCTGATCGGCGCCACCGCGTGAACATCATCGGCGCGCAACATGTGTGACCGCAGCTGGG 656  
Db 1731 CGTGTGTGTGGGCGCCACACCGCTCAACATCATCGGAGCGCAACCTGTTGAGCAGAGATCGG 1790  
Qy 657 CTGCACCCCTGAATTCCTCCCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAGCCCGG 716  
Db 1791 TTGCACGCTGAATTCCTCCCATTAGCCCTATCGAGCGGTACCGGTGAAGCTGAAGCCCGG 1850  
Qy 717 CATGAGCGGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGAC 776  
Db 1851 GATGGACGGCGCGAAGGTCAAGCAATGGCCATTTGACAGAGGAGAGATCAAGGACCTGGT 1910





QY 129 CCGCAACTGCGCGCCCGCCCGCCCGCAAGAGGGCTGTGGAAAGTGCAGCAAGGAGGGCCACCA 188  
Db 1541 CCGCAACTGCGAGGGCCCGCCCGCCCGCCCGCAAGAGGGCTGTGGAAATGTGGAAGGAGGACACCA 1600  
QY 189 GATGAAGGACTGCAACCGAGCGCCAGGCCCAACTTCTTCCGCGAGGACTGGGCTTCCCGCCA 248  
Db 1601 AATGAAAGATTGTACTGAGAGACAGGCTAA-FTTTTATAGGGAAGATCTGGGCTTCCCGCA 1659  
QY 249 GGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAAACCGCGCCCAACAGCCCGCAACAGCGCGGA 308  
Db 1660 AGGGAAGGCCAGGGAATTTCTTCAAGAGACAGCAGAGCCACAGCCCGCAACAGAGAGAGA 1719  
QY 309 GCTGAGGTGCGCGG-----CGAACACCCCGCAGCGAGCGCGCGCGCGAGCGCGAGGG 362  
Db 1720 GCTTCAGGTTTGGGGAAGAGAGCAACAACTCCCTCTCAGAAAGCAGGAGCCGATAGACAAGG 1779  
QY 363 CA-----CCCTGAACCTTCCCGCAGATCAACCTGTGGCAGCGCCCGCTGTGAGCATCA 416  
Db 1780 AACTGTATCTTTCAGTCTTCCCTCAGATCACTTCTTGGCAGCGACCCCTCGTCACAATAA 1839  
QY 417 GGTGGCGCGCAGATCAAGAGGGCCCTGTCTGACACACCGCGCGCAGACACACCGTGTGGA 476  
Db 1840 GATAGGGGGCAGCTCAAGAGGCTCTCTTGGACACCGGAGCAGAGACACCGTGTGGA 1899  
QY 477 GGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGGCAATCGCGGCTTCAT 536  
Db 1900 GGAGATGTCGTTGCGAGCGCGCTGGAAGCCGAAGATGATCGGGGAATCGCGGTTTCAT 1959  
QY 537 CAAGGTGCGCAGTACGACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCAC 596  
Db 1960 CAAGGTGCGCAGTATGACAGATCTCTCATCGAAATCTGCGGCCCAAGGCTATCGGTAC 2019  
QY 597 CGTGTGATCGGCGCCCGCCCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGG 656  
Db 2020 CGTGTGTGGCGCCCGCCCGTCAACATCATCGGACGCAACCTGTTGACGCGAGATCGG 2079  
QY 657 CTGCAACCTGAATTCCTCCATCAGCCCGCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGG 716  
Db 2080 TTGACGCTGAATTCCTCCATTAGCCCTATCGAGAGCGTACCGGTGAAGCTGAAGCCCGG 2139  
QY 717 CATGACCGCCCGCAAGGTGAAGCATGTGCCCTGTGACCGAGAGAGATCAAGGCCCTGAC 776  
Db 2140 GATGACCGCCCGCAAGGTCAAGCAATGGCCATTTGACAGAGGAGAGATCAAGGCACTGGT 2199  
QY 777 CGCCATCTGCGAGAGATGGAAGAGGAGGCGCAAGATCAACCAAGATCGSCCCCGAGAACCC 836  
Db 2200 GGAGTTTGCAGAGATGGAAGAGGAGGAAATCTCCAAGATTGGGCTTGAGAACCC 2259  
QY 837 CTACAACACCCCGTGTTCGCCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGT 896  
Db 2260 GTACAACACCGCGTGTTCGCAATCAAGAAAGAGGACTCGACGAAATGGCGCAAGCTGGT 2319  
QY 897 GGACTTTCGCGAGCTGAACAGCGGACCCAGGACTTCTGGAGGTGACGTGGGCAATCCC 956  
Db 2320 GGACTTTCGCGAGCTGAACAGCGGACCCAGGACTTCTGGAGGTTCAGCTGGGCAATCCC 2379  
QY 957 CCACCCCGCGGCTGGAAGAGAGAGAGCGTGCAGCTGTGACGTGGCGGAGCGCTA 1016  
Db 2380 GCACCCCGCGGCTGGAAGAGAGAAATCTCGTACCGTACTGTGATGTGGGTGATGCTA 2439  
QY 1017 CTTGAGGTGCGCTTGGACGAGGACTTCGCGAAGTACACCGCTTTCACCATCCCGCAGCAT 1076  
Db 2440 CTTCTCGTTCCTTGGACGAGGACTTCAGAGAGTACACTGCTTCACATCCCTTCGAT 2499  
QY 1077 CAACAAAGAGACCCCGGCAATCGCTACAGTACAAAGTGTGCGCCAGGCTGGAAGGG 1136  
Db 2500 CAACAAAGAGACACCGGGGATTCGATATCAGTACAAAGTGTGCGCCAGGCTGGAAGG 2559  
QY 1137 CAGCCCCAGCATCTTCAGAGCAGCATGACCAAGATCTTGAGGCGCTTCCGCGCCCGCA 1196  
Db 2560 CTCTCCCGCAATCTTCAGAGTAGATGACCAAAATCTCTGAGCTTTTCGCGCAACAGAA 2619

QY 1197 CCCGAGATCTGTGATCTTACCA-----GGCCCCCTGTACGTGGGCAAGCAGCTCGAGAT 1250  
Db 2620 CCCGACATCTGATCTATCATGTAGTACATGTGTACGTGGGCTCTGTACTAGAT 2679  
QY 1251 CGGCGACACCGCGCAAGATCGAGGAGCTGGCAAGCAGCTGTGCGCTGGGCTTCAC 1310  
Db 2680 AGGGGAGCACCAGCAGATCGAGGAGCTGGCCAGCAGCTGTGTGAGGTGGGAGTAC 2739  
QY 1311 CACCCCGCACAAGAGACACCAAGAGAGGCCCGCTTCTTCTGCCCCT-----CGAGCTGCA 1364  
Db 2740 CACACCGCACAAGAGACACCAAGAGAGGCTTCTTCTTCTGATGGGTAGGAGCTGCA 2799  
QY 1365 CCGCGACAAGGTGAGACCGTGCAGCCCATCGAGTGCCTCGAGAAAGAGAGCTGGAACGTA 1424  
Db 2800 CCCTGACAAATGAGACCGGTGCAGCCCTATCGTGTGCGAGAAAGACAGCTGAGCTGTCAA 2859  
QY 1425 CGACATCCAGAGCTGTGGGCAAGCTGAACTGGGGCCAGAGCTTACCCCGGATCAA 1484  
Db 2860 CGACATACAGAACTGTGGGGAAGTTGAATGGGCCAGTCAAGATTTTACCACGAGGATTA 2919  
QY 1485 GGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCCGACATCTGCCCCCT 1544  
Db 2920 GGTGAGGCGAGCTGTGCAAACTCTTCCGCGGAACCAAGGCACTCACAGAGGTGATCCCCCT 2979  
QY 1545 GACCGAGAGGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTTGCOCGAGCCCGTGA 1604  
Db 2980 AACCGAGAGGCCGAGCTCGAACTGGCAGAAAAACGAGAGATCTTAAAGGAGCCCGTGA 3039  
QY 1605 CGGCGTGTACTACACCCCGCAGCAAGGACCTGTGGTGGCGAGATCCAGAGCAGGGCCACA 1664  
Db 3040 CGGCGTGTACTATACCCCTTCAAGGACCTGATCGCCGAGATCCAGAGCAGGGGCAAGG 3099  
QY 1665 CCAGTGAACCTTACAGATCTACAGAGGACCTTCAAGAACCTGAAAGCCGCAAGTACGC 1724  
Db 3100 CCAGTGAACCTTACAGATTTTACAGAGGACCTTCAAGAACCTGAAAGCCGCGAGTACGC 3159  
QY 1725 CAAGATCGCACCGCCCAACCAACGACGCTGGAAGCAGCTGACCGAGGCGCTGCAGAAAT 1784  
Db 3160 CCGGATGAGGGGTGCCACACTAAGCAGCTCAAGCAGCTGACCGAGGCGCTGCAGAAAT 3219  
QY 1785 CGCCATCGAGAGCATCTGTGATCTGGGGCAAGACCCCGCAAGTTCCGCTGCCCTCCAGAA 1844  
Db 3220 CACCAACGAAAGCATCTGTGATCTGGGGAAGAGCTCTTAAAGTTCAAGCTGCCCTCCAGAA 3279  
QY 1845 GGAGACCTTGGGAGACCTGTGGACCGACTCTGCGAGCCACCTGGAATCCCGGAGTGGGA 1904  
Db 3280 GGAACCTTGGAAACCTGTGGGACAGAGATTGGCAGGCCACCTGGATTCCTGTAGTGGGA 3339  
QY 1905 GTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTATACAGCTGGAGAGGAGCCCATCAT 1964  
Db 3340 GTTCGTCAACACCCCTTCCCTGTGTGAAGCTGTGTATCCAGCTGGAGAGGAGCCCATAGT 3399  
QY 1965 CGGCGCGAGACCTTCTGATGTGAGCGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2024  
Db 3400 GGGCGCGCAACCTTCTACGTGATGGGGCGCGCTTAAAGGGAGACTAAGCTGGGCAAGC 3459  
QY 2025 CGGCTAGCTGACACCGCGGCGCGCAGAGATCGTGAAGCTGACCGGAGACCAACCA 2084  
Db 3460 CGGATAGCTCACTAACCGGGGACAGAGAGGTTGTCACTCTGAGACCAACCA 3519  
QY 2085 GAAGACGAGCTGAGGCCCATCCAGCTGGCGCTTCCAGAGCAGCGCGAGGAGTGAACAT 2144  
Db 3520 GAAGACTGAGCTGAGGCCCATTTACCTCGCTTTCAGGACTCGGGCTTGGAGGTGAACAT 3579  
QY 2145 GTGACCGCAGCAGCTGAGCGCTTGGGATCATCTCAGGCCAGCCCGACAAAGAGCGAGAG 2204  
Db 3580 CGTGACAGACTCTCAGTATGCTTGGGATCATTTCAAGCCCGCAGCCAGACCAAGAGTGTG 3639  
QY 2205 CGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTG 2264  
Db 3640 CGAGCTGTCAATCAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTATCTATGCGCTG 3699  
QY 2265 GGTGCGCGCCCAAGGGGCTCGCGGCGCAACGAGCAGATCGAACAGCTGTGTGAGCAAGGG 2324



Db 3700 GTATCCCGCCCAAGGCAATGGCGCAATGAGCAGGTGCGAAGCTGGTCTCGGCTGG 3759  
QY 2325 CATCGCAAGGTGCTGTTCTCTGGACGGGATCGA 2357  
Db 3760 CATCAGGAAGGTGCTATTCTCGATGGCATCGA 3792

RESULT 12  
US-09-552-950-5  
; Sequence 5, Application US/09552950  
; Patent No. 6541248  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; TITLE OF INVENTION: Anti-Viral Vectors  
; FILE REFERENCE: 674524-2004  
; CURRENT APPLICATION NUMBER: US/09/552,950  
; CURRENT FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 9772  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: pSYNGP  
US-09-552-950-5

Query Match 65.1%; Score 1599.4; DB 4; Length 9772;  
Best Local Similarity 81.5%; Pred. No. 1e-263;  
Matches 1934; Conservative 0; Mismatches 411; Indels 28; Gaps 6;

QY 12 CATGCCGAGGCGCATGAGCCAGG----CCACAGGCGCAACATCTGTATGTCAGCGCAGCA 68  
Db 2193 CCGTGTGAGGCGCATGAGCGAGTGAACCACTCCGCTACCATCATGTCAGCGCGGCA 2252  
QY 69 CTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGCAAGGAGGCGCATCGC 128  
Db 2253 CTTTCGGAACCAACGCAAGATCGTCAAGTCTTCAACTGTGCAAGAGGCGCACACAGC 2312  
QY 129 CGCAACTGCGCGCGCCCGCAAGAGGCGTGTGGAAGTGGCGCAAGGAGGCGCACCA 188  
Db 2313 CGCAACTGCGAGCGCCCTTAGGAAAAGGCGTGTGGAATGTGGAAGAGGACCA 2372  
QY 189 GATGAAGACTGACCGAGCGCGCAGCCAACTTCTTCGCGAGGACCTGGCCCTTCCCGCA 248  
Db 2373 AATGAAGATTGTACTGAGAGACAGGCTAA-TTTTGTAGGGAAGATCTGGCCCTTCCCA 2431  
QY 249 GGGCAAGCGCCGAGTTCCTCCAGCGAGCAGAACCGCGCAACAGCCCGCAGCGCGGA 308  
Db 2432 AGGGAAGCGCCAGGGAATTTCTTCAGAGCAGACCAAGCCCAAGCCCGCAGAGAGA 2491  
QY 309 GCTCAGGTGCGCGG-----CGACAACCCCGCAGCGAGGCGCGCGCGCAGCGCAGGG 362  
Db 2492 GCTTCAGGTTTGGGAAGAGACAACAACCTCCCTCTCAGAAGCAGGAGCGATAGACAAG 2551  
QY 363 CA-----CCCTGAATCTCCCGCAGATCAACCTGTGGCAGCGCCCTCTGGTGAAGATCAA 416  
Db 2552 AACTGTATCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAATAA 2611  
QY 417 GGTGGCGCGCAGATCAAGGAGCGCTGTGACACCGCGCGCAGCAGACCGCTGCTGA 476  
Db 2612 GATAGGGGGAGCTCAAGAGGCTCTCTTGACACCGGAGCAGACGACCGTCTGA 2671  
QY 477 GGAGATGAGCTCGCCCGCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCAT 536  
Db 2672 GGAGATGCTGTTGCCAGGCGCTGGAAGCCGAAGATGATCGGCGGAATCGCGGTTTCAT 2731  
QY 537 CAAAGTGCAGATGACGACAGATCTGTATGATGATCTCGGCGCAAGAGGCCATCGGCAC 596  
Db 2732 CAAAGTGCAGATGACGACAGATCTGTATGATGATCTCGGCGCAAGAGGCCATCGGTAC 2791  
QY 597 COTGCTGATCGGCGCCCGCCCGTGAACATCATCTCGGCGCAACATGCTGACCCAGCTGG 656

Db 2792 CGTGTGTGGGCCCCCAACCCGTCACATCATCGGACGCAACCTGTTGACGACGATCGG 2851  
QY 657 GTGCAACCTGTAACTTTCCTCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGG 716  
Db 2852 TTGCACGCTGAACCTTTCCTCCATTTAGCCCTATCGAGACGGTACCGGTGAAGCTGAAGCCCGG 2911  
QY 717 CATGACGCGCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAGATCAAGGCCCTGAC 776  
Db 2912 GATGACGCGCCCGAAGGTCAAGCAATGGCCATTGACAGAGGAGAGATCAAGGCACTGT 2971  
QY 777 CGCCATCTGCGAGGAGATGAGAGAGGAGGCAAGATCAACAGATCGGCCCGCAGAACCC 836  
Db 2972 GGAGATTTCACAGAGATGGAAGAGGAGGAATCTCCAAGATTGGCCCTCGAGAACCC 3031  
QY 837 CTACAACACCCCGTGTTCGCCATCAAGAAAGAGACAGCAACAGTGGCGGCAAGCTGT 896  
Db 3032 GTACAACACGCGCGTGTTCGCAATCAAGAAAGAGGACTCGACGAATGGCGAAGCTGT 3091  
QY 897 GGACTTTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGACAGCTGGGATCC 956  
Db 3092 GGACTTTCGCGAGCTGAACAGCGCACCGCAAGACTTCTGGGAGGTTCAGCTGGGATCC 3151  
QY 957 CCACCCCGCGCTGAAAGAAAGAGAGGCTGACCGTGTGTGACGTGGGCGACGCTTA 1016  
Db 3152 GCACCCCGCGGCTGAAGAGAGAAATCGTGACCGTACTGGATGTGGGTGATGCTTA 3211  
QY 1017 CTTAGCTGCTGCGAGGAGCTTCGCAAGATGACACCGCTTCAACATCCCGAGAT 1076  
Db 3212 CTTCTCCGTTCCCTTGGAGGAGCTTCAGGAAAGTACATGCTGCTTCAACATCCCTTCGAT 3271  
QY 1077 CAACAAACAGACCCCGCATCCGCTACAGTACAACTGTGAGGCGCTTCCCGCGCCGCAA 1136  
Db 3272 CAACAAACAGACACCGGGATTCGATATCAGTACAACTGTGCTGCCCGAGGCTGGAAGG 3331  
QY 1137 CAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCTTGAGGCGCTTCCCGCGCCGCAA 1196  
Db 3332 CTCTCCCGCAATCTTCAGAGTAGCATGACCAAAATCTTGAGGCTTTCGCAACAGAA 3391  
QY 1197 CCGCGAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGCGGACCTGAGAT 1250  
Db 3392 CCGCGACATCTGTATCTATCAGTATCAGTATGATGATGATGATGATGATGATGATGAT 3451  
QY 1251 CGGCGAGCAGCGCGCAAGATCGAGGAGTGGCGAAGCAGCTGTGCGTGGGCTTCA 1310  
Db 3452 AGGCGAGCAGCGCACCAAGATCGAGGAGTGGCGCAGCAGCTGTGAGGTGGGAGTAC 3511  
QY 1311 CACCCCGCACAAGAGCAGCAGAGAGGAGCGCCCTTCTGCGCCAT-----CGAGCTGA 1364  
Db 3512 CACACCCGCAAGAGCAGCAGAGAGGAGCTTCCCTTCTCTGATGGGTTACGAGCTGA 3571  
QY 1365 CCGCGAAGTGGACCGTGGAGCCCATCGAGTGGCCGAGAGAGAGAGTGGACCGTGA 1424  
Db 3572 CCGTGAACATGGACCGTGGAGCTTATGCTGTGTCAGAGAAAGACAGCTGAGCTGCA 3631  
QY 1425 CGACATCCAGAGCTGGTGGCGAAGCTGAACCTGGGCGCAGCAGATCTACCCCGGATCAA 1484  
Db 3632 CGACATAAGAGCTGGTGGGGAAGTTGAACCTGGGCGAGTCHAGATTTACCCAGGATTA 3691  
QY 1485 GGTGGCGCAGCTGTCAAGAGCTGCTGCGGGGGCGCAAGCGCCCTGACCGCAGCATCTG 1544  
Db 3692 GGTGAGGCGAGTGTGCAAACTCTCCGCGGAACCAAGGCACTCAAGAGGTGATCC 3751  
QY 1545 GACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGGCGGAGCCGTGA 1604  
Db 3752 AACCGAGGAGCGGAGCTCGAAGCTGGCAGAAAACCGAGAGATCTCTAAAGGAGCC 3811  
QY 1605 CGGCGTGTACTACGACCCCGCAAGGAGCTGTGCGCAGAGTCCAGAGAGCAGGCGCA 1664  
Db 3812 CGGCGTGTACTATGACCCCTCCAGAGACTGATCGCGAGATCCAGAGAGGCGGAGG 3871  
QY 1665 CAGGTGGAGCTTACAGATCTTACAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTAC 1724

Db 3872 CCAGTGGACCTATCAGATTACAGAGGCCCTTCAAGAACTTGAAGACCGGCAAGTACGC 3931  
Qy 1725 CAAGATGGCCACCGCCACCAACAGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAAAT 1784  
Db 3932 CGGATGAGGGGTGCCCCACACTAACAGCTCAAGCAGCTGACCGAGGCGGTGCAGAAAT 3991  
Qy 1785 CGCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATTCAGAA 1844  
Db 3992 CACCACCGAAAGCATCGTGATCTGGGGAAAGACTCTTAAGTTCAAGCTGCCCATTCAGAA 4051  
Qy 1845 GGAGACCTGGGAGACCTCGTGACCGACTACTTGGCAGGCGCCACTGGATGCCGAGTGGGA 1904  
Db 4052 GGAACCTTGGGAACCTGGTGACAGATATTGGCAGGCGCACTTGGATTCCTGAGTGGGA 4111  
Qy 1905 GTTCGTGAACACCCCGCCCTCGTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCAT 1964  
Db 4112 GTTCGTGAACACCCCGCCCTCGTGGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATAGT 4171  
Qy 1965 CGGCGCCGAGACCTTCTAGTGGAGCGGCGCGCCGACCAACCGCAGAGCAAGATCGGGCAAGGC 2024  
Db 4172 GGGCGCCGAAACCTTCTACGTGGATGGGGCGGCTAACAGGGAGACTAAGCTGGGCAAGC 4231  
Qy 2025 CGGCTACGTGACCGAGCGGCGCGCAGAAAGATCGTGAGCCTGACCGAGAGCAAGATCGGGCAAGGC 2084  
Db 4232 CGGATACGTCTACTAACCGGGGCGAGACAGAGGTTGTACCCCTCACTGACACCAACCA 4291  
Qy 2085 GAAGACCGAGCTGCAGGCGCATCCAGCTGGCGCTGCAGGACAGCGCGAGAGGTGAACAT 2144  
Db 4292 GAAGACTGAGCTGCAGGCGCATTTACCTCGCTTTCAGAGACTCGGGCTGGAGGTGAACAT 4351  
Qy 2145 CGTGACGACAGCGAGTACGCGCTTGGGCGCATCATCCAGGCCCGAGCGCAAGAGCGAGAG 2204  
Db 4352 CGTGACGACACTCTCAGTATGCGCTGGGCGCATCATTCAGGCCCGAGCGAGAGTGAGTC 4411  
Qy 2205 CGAGCTGGTGAACAGACATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACTTGAGCTG 2264  
Db 4412 CGAGCTGGTCAATCAGATCATCGAGCAGCTGATCAAGAGGAGAAAGGTCTATCTGGGCTG 4471  
Qy 2265 GGTGCGCGCCCAAGGGCATCGCGGCAAGCAGAGCAGATCGACAGCTGGTGAGCAAGGG 2324  
Db 4472 GGTACCGCGCCCAAGAGGCAATGGCGCAATGAGCAGGTGCAAGCTGTCTCGGCTGG 4531  
Qy 2325 CATCGCAAGTGTCTTCTGGAGCGGCATCGA 2357  
Db 4532 CATCAGGAAGTGCTATTCTCTGGATGGCATCGA 4564

## RESULT 13

US-09-872-733A-6  
; Sequence 6, Application US/09872733A  
; Patent No. 6656706  
; GENERAL INFORMATION:  
; APPLICANT: The Government of the United States of America, as  
; TITLE OF INVENTION: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND  
; TITLE OF INVENTION: SIV ENV GENES  
; FILE REFERENCE: 2026-4287US1 HIV GAG/POL, SIV GAG & ENV  
; CURRENT APPLICATION NUMBER: US/09/872,733A  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: PCT/US00/34985  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/173,036  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 8366  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: DNA sequence  
; OTHER INFORMATION: of the construct pcwvgagpolBnkan containing a CMV  
; OTHER INFORMATION: promoter, a HIV gag/pol gene and a kanamycin  
; OTHER INFORMATION: resistance gene

## US-09-872-733A-6

Query Match 62.6%; Score 1539.2; DB 4; Length 8366;  
Best Local Similarity 79.9%; Pred. No. 1.8e-253;  
Matches 1896; Conservative 0; Mismatches 448; Indels 28; Gaps 6;  
Qy 14 TGGCGAGGCGCATGAGCGAGGCCCAACAGC---GCCAACATCTCTGATGAGCGCAGCAACT 70  
Db 1857 TGGCGAGGCGCATGAGCGAGGTCAGAACTCGGCGACCATTAATGATGAGAGAGGCAACT 1916  
Qy 71 TCAGAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGGCACATCGCCC 130  
Db 1917 TCCGGAACCCAGCGGAAGATCGTCAAGTCTTCAATTTGTGGCAAGAGGCGCACACCGCA 1976  
Qy 131 GCAACTGCGCGCGCCCGCCGCAAGAGGCGCTCTCGAAGTGGCGAAGAGGCGCCACACAGA 190  
Db 1977 GGAATCTCGCGCGCCCGCCGCAAGAGGCGCTGTGGAAATGTGGAAAGAGAGACACCAA 2036  
Qy 191 TGAAGGACTGCAACGAGCGCCAGGCCCAACTTTCTTCGCGAGGAGCTGCGCTTCCCGCAGG 250  
Db 2037 TGAAGGATTTGATCTGAGAGACAGGCTAA-TTTTTTAGGGAAGATCTGGCGCTTCTTACAAG 2095  
Qy 251 GCAGAGCGCGCGAGTTCCCGAGCGAGAGACCGGCGCAACAGCCCGCCACAGCGCGCAGC 310  
Db 2096 GGAAGGCGCAGGGAATTTTCTTCAGAGCAGACCGAGCGCAACAGCCCGCCACAGAGAGAGC 2155  
Qy 311 TGCAGGTGCGCGG-----CGAACACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGAGGCA 364  
Db 2156 TTCAAGTCTGGGTAGAGACAACTCTCCCTCAGAAAGCAGAGGCGCGATAGACAAAGAA 2215  
Qy 365 -----CCCTGAACTTCCCGCAGATCACTCTGTGGCAGCGCCCTTGTGTAGCATCAAGG 418  
Db 2216 CTGTATCTTTAACTTCTCCCTCAGATCACTCTTTGGCAACGACCCCTCGTCACAGTAAGA 2275  
Qy 419 TGGCGCGCGAGATCAAGAGGCGCTCTGTGACACCGGCGCGCGAGCAACCGTGTGTGAGG 478  
Db 2276 TCGGGGGGCAACTCAAGGAGCGCTGTCTCGATACAGGAGCAGATGATACAGTATTAGAG 2335  
Qy 479 AGATGAGCTGCGCGGCAAGTGAAGCGCAAGATGATCGGCGGATCGCGCGCTTTCATCA 538  
Db 2336 AATGAGTTTGGCAGGAAGATGGAACCAAAATGATAGGGGGATCGGGGGCTTTCATCA 2395  
Qy 539 AGTGCGCCAGTACGACAGATCTCTGATCGAGATCTCGGCAAGAGGCCATCGGCACCG 598  
Db 2396 AGGTGAGCGAGTACGACAGATCTCATAGAAATCTGTGGACATAAAGCTATAGGTACAG 2455  
Qy 599 TGCTGATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCT 658  
Db 2456 TATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTGACCCAGATCGGCT 2515  
Qy 659 GCACCTGAACTTCCCGCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCGCCGCA 718  
Db 2516 GCACCTTGAATTTCCCGCATCAGCCCTATTGAGACGGTCCCGCTGAAGTTGAAGCGGGGA 2575  
Qy 719 TGAAGCGCCCAAGGTGAAGCAGTGGCCCTCGACCGGAGGAGAGATCAAGGCCCTTGACCG 778  
Db 2576 TGAAGCGCCCAAGGTCAAGCAATGGCCATTGACGAAAGAGAGATCAAGGCCCTTAGTCG 2635  
Qy 779 CCATCTCGAGGAGATGGAGAGGAGGCGAAGATCACCAAGATCGGCGCCCGAGAACCCCT 838  
Db 2636 AAATCTGTACAGAGATGGAGAGGAGGAGAGATCAGCAAGATCGGGCTTGAGAACCCCT 2695  
Qy 839 ACAACACCCCGTTCGCGCATCAAGAGAGGAGGAGCAGCACCAAGTGGCGCAAGCTGGTG 898  
Db 2696 ACAACACTCCAGTCTTCGCAATCAAGAGAGGAGGAGTACCAAGTGGAGAAAGCTGGTG 2755  
Qy 899 ACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCGATCCCC 958  
Db 2756 ACTTCAGAGAGCTGAACAAGAGAACTCAGGACTTCTGGGAAGTTTCAGCTGGGCATCCAC 2815  
Qy 959 ACCCGCGCGCTGAAGAGAGAGAGAGCGGTGACCGCTGTGGAGCTGGGCGAGCGCTACT 1018  
Db 2816 ATCCCGCTGGGTTGAAGAGAGAAAGTCAAGTGAAGTGTGGTGTGGTGTGGTGTGGTGTACT 2875



Db 1384 TTCAGGCTGGGTTAGACAAACAACCTCCCTCAGAAAGCAGGAGCCGATAGACAAGGAA 1443  
Qy 365 -----CCCTGAACCTCCCCAGATCAACCTGTGTGCGAGCGCCCTGTGTGAGCATCAAGG 418  
Db 1444 CTGTATCCTTTAACTTCCCTCAGATCACTTTTGGCAAGACCCCTCTGTACAGTAAGGA 1503  
Qy 419 TGGGCGGCGAGATCAAGAGGCGCTGTGTGGAACCGCGCGCGACACACCGTGTGTGAGG 478  
Db 1504 TCGGGGGGCAACTCAAGGAAGCGCTGTGTGATACAGGAGCAGATGATACAGTATTAGAA 1563  
Qy 479 AGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGGGGGATCGCGGCTTCATCA 538  
Db 1564 AAATGAGTTTGCAGGAAGATGGAACCAAAATGATAGGGGGATCGGGGGCTTCATCA 1623  
Qy 539 AGGTGCGCAGTACGACAGATCTGTGATCGAGATCTGCGGCAAGAGCCCATCGGCAAG 598  
Db 1624 AGGTGAGCAGTACGACAGATCTATAGAAATCTGTGGAATTAAGCTATAGGTACAG 1683  
Qy 599 TGCTGATCGGCGCCNACC-----CCGTGAACATCATCGGCGCGAATCTGCTGACCCAGCTG 654  
Db 1684 TATTAGTAGGACCTACCTTACACCTGTCAACATAATTGGAAGAAATCTGTTGACCCAGATC 1743  
Qy 655 GGCTGCAACCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCC 714  
Db 1744 GGCTGCACTTGAACCTTCCCATCAGCCCTATTGAGACGGTGCCTGGAAGTGAAGCCG 1803  
Qy 715 GGCTGAGCGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTG 774  
Db 1804 GGGATGAGCGGCCCAAGGTCAAGCAATGGCCATTGACAAAGAGAGATCAAGGCCCTTA 1863  
Qy 775 ACCGCGATCTCGGAGGAGATGAGAGAGGAGGCAAGATCAACAAGATCGGCCCGAGAAC 834  
Db 1864 GTCGAAATCTGTACAGAGATGAGAGAGGAGGAGAGATCAGCAAGATCGGCCCTGAGAAC 1923  
Qy 835 CCCTACAAACCCCGCTGTCCCATCAAGAGAGGAGACAGACCAAGTGGCGCAAGCTG 894  
Db 1924 CCCTACAAACTCCAGTCTTCCCATCAAGAGAGGAGACAGTACCAAGTGGAGAAAGCTG 1983  
Qy 895 GTGACCTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATC 954  
Db 1984 GTGACCTTCCGAGAGCTGAACAAGAGACTCAGGACTTCTGGGAGTTTCACTGGGCATC 2043  
Qy 955 CCCACCCCGCGGCTGAAGAAGAGAGAGCGTGCCTGTGTGAGCTGGGCGAGCGCC 1014  
Db 2044 CCACATCCCGCTGGGTTGAAGAAGAGAGTCAAGTCAAGTGTGGATGTGGGTGATGCC 2103  
Qy 1015 TACTTCAGCTGCGCTCGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGC 1074  
Db 2104 TACTTCTCCGTTCCCTTGAAGAGAGACTTCAAGGAGTACACTGCTTCCACGATACCTAGC 2163  
Qy 1075 ATCAACACGAGACCCCGGCTCCGCTACCAAGTACCAAGTGTGCTGCCCGAGGCTGGAAG 1134  
Db 2164 ATCAACACGAGACCCCGGCTCCGCTACCAAGTGTGCTGCCACAGGATGGAAG 2223  
Qy 1135 GGCAGCCCGAGATCTTCCAGAGCAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGC 1194  
Db 2224 GGATCAGCAGCATCTTCAAGCAGCAGATGACCAAGATCTGTGAGCGCTTCCGCAAGCAA 2283  
Qy 1195 AACCOCGAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGCGAGCGACTGGAG 1248  
Db 2284 AACCOCGAGATCTGTATCTATCTAGTACATGACGACCTCTACGTAGGAAGTGAACCTGGAG 2343  
Qy 1249 ATC-GGGCAGCAGCGCGCAAGATCGAGGAGCTCGCAAGCACTGTGCTGCGCTGGGGCTT 1307  
Db 2344 ATCGGGCAGCAGCAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGTGGGACT 2403  
Qy 1308 CACACCCCGAGCAAGAGCAACAGAGGAGCCCTTCTTCTGCCCAT-----CGAGCT 1361  
Db 2404 GACCACACCAAGAGCAACAGAGCAACAGAGGAGCACTTCTGCTGAGTGGGCTACGAACT 2463  
Qy 1362 GCACCCCGAGCAAGTGGAGCGCTGAGCGCCATCGAGCTGCGCGAGAGAGAGCTGGACGT 1421  
Db 2464 GCATCTCTGACAAAGTGGAGCAGTGGAGCGCCATCTGCTGCTGAGAGGAGCAGCTGGACTGT 2523

Qy 1422 GAACGACATCAGAAAGCTGTGGCAAGCTGAACTGGGCGCAGCCAGATCTTACCCGGCAT 1481  
Db 2524 GAACGACATCAGAAAGCTGTGGCAAGCTGTGAACTGGGCAAGCCAGATCTTACCCAGCAT 2583  
Qy 1482 CAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGGCGCCAAAGGCGCTGACCGACATCGTGCC 1541  
Db 2584 CAAGTGTAGGCGCTGTGCAAGCTGTTCGAGGAAACCAAGGCACTGACAGAAAGTATGCC 2643  
Qy 1542 CTTGACGAGGAGCGCAGCTGTGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCCCGT 1601  
Db 2644 ACTGACGAGGAGGAGCTGTGAACTGGCAGAGAACCGAGAGATCTCTGAAAGAGCCAGT 2703  
Qy 1602 GCACGGCGTGTACTACGACCCAGCAAGCACTGTGTGGCGGAGATCTCAGAAAGCAGGCCA 1661  
Db 2704 ACATGGAGTGTACTACGACCCAGCAAGCACTGTATGCGCAGAGATCCAGAGCAGGGCA 2763  
Qy 1662 CGACAGTGTGACCTACCGATCTACAGGAGCCCTTCAAGAACTGTGAAGACCGGCAAGTA 1721  
Db 2764 AGGCAATGGACCTTACCAATCTACAGGAGCCCTTCAAGAACTGTGAAGACAGGCAAGTA 2823  
Qy 1722 CGCAGATGCGCACCGCCACACCAAGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAA 1781  
Db 2824 CGCAAGATGAGGCTGCGCCACCAAGATGTGAAGCAGCTGACAGAGGCGAGTGCAGAA 2883  
Qy 1782 GATCGCATGAGAGCAGCTGTGATCTGGGCAAGACCCCAAGTTTCCGCTGCCCATCCA 1841  
Db 2884 GATCACCACAGAGAGCATGTGATCTGGGCAAGACTTCCCAAGTTTCAAGCTGCCCATACA 2943  
Qy 1842 GAAGGAGACTTGGGAGACCTGTGTGAGCCGACTATGTGGCAGGCGACCTGGATCCCGAGTG 1901  
Db 2944 GAAGGAGACTTGGGAGACATGTGTGAGCCGAGTACTGTGCAAGCCACTGGATCCCTGAGTG 3003  
Qy 1902 GGAGTTCGTGAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGAGAGGAGGAGCCCAT 1961  
Db 3004 GGAGTTCGTGAACACCCCGCTGTGTGAGCTGTGTGAGCTGTGAGAGGAGGAGCCCAT 3063  
Qy 1962 CATCGCGCGCAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAA 2021  
Db 3064 CGTGGAGCAGAGACCTTCTACGTGGATGGGCGAGCCCAACAGGAGACCAAGCTGGGCAA 3123  
Qy 2022 GCGCGCTAGCTGACGACCGGCGCGCGCAGAGATCTGTGAGCTGTGACCGAGACCAACAA 2081  
Db 3124 GCGAGGCTAGCTGACCAACCGGAGCAGCAGAAAGTGTGACCTGACTGACACCAACCAA 3183  
Qy 2082 CCAGAACCGAGCTGAGGCGCATCTGAGCTGGCGCTGCAGACAGCGGCGAGGAGTGA 2141  
Db 3184 CCAGAACGAGCTGAGGCTGCAAGCCATCTACCTAGCTTGTGCAAGCAGCGGACTGGAGTGA 3243  
Qy 2142 CATGTCAGCAGCAGCAGCTACGCTGGGCGATCATCCAGGCCAGCCCGCAAGAGCGA 2201  
Db 3244 CATGTCAGAGCTCAGTACGCTAGC-CATGGGCGATCATCCAGCACAACAGACCAATCCGA 3302  
Qy 2202 GAGCAGCTGTGAACAGATCTATGAGCAGCTGATCAAGAGGAGAGAGTGTACTTGAG 2261  
Db 3303 GTCAGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTTGAG 3362  
Qy 2262 CTGGGTCCCGCCACAGGCGCATCGGCGCAAGCAGCAGATCCAGAGCTGTGTGAGCAA 2321  
Db 3363 ATGGGTACCAAGCAGCAAGAGGAAATTTGAGAGGAAATGAACAGTAGATAAATAGTCAAGTGC 3422  
Qy 2322 GGGCATCCGCAAGGCTGTCTTCTTGGAGCGGCTCATCGAT 2358  
Db 3423 TGGGATCCGGAAGGCTGTCTTCTTGGAGCGGATCGAT 3459

RESULT 15  
US-09-952-060-1  
; Sequence 1, Application US/09952060  
; Patent No. 673993  
; GENERAL INFORMATION:  
; APPLICANT: Emini, Emilio A.  
; APPLICANT: Youil, Rima

```

; APPLICANT: Bett, Andrew J.
; APPLICANT: Chen, Ling
; APPLICANT: Kaslow, David C.
; APPLICANT: Shiver, John W.
; APPLICANT: Toner, Timothy J.
; APPLICANT: Casimiro, Danilo R.
; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND
; TITLE OF INVENTION: MODIFICATIONS
; FILE REFERENCE: 20747Y
; CURRENT APPLICATION NUMBER: US/09/952,060
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: PCT/US01/28861
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/317,814
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 60/279,056
; PRIOR FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/233,180
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2577
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Codon optimized DNA encoding modified wt pol
US-09-952-060-1

Query Match          52.2%; Score 1282.8; DB 4; Length 2577;
Best Local Similarity 85.6%; Pred. No. 7.3e-210; Mismatches 232; Indels 12; Gaps 2;
Matches 1454; Conservative 0;

QY 672 CCCCATCAGCCCATCGAGACCGTGCCTGAAAGCTGAAAGCCCGGATGAGCGGCCCAA 731
DB 15 CCCCCTCTCCCATTTGAGACTGTGCTGTGAAGCTGAAGCTGAAGCTGGCATGGATGGCCCCAA 74

QY 732 GGTGAAGCAGTGGCCCTGACCGAGGAGAAAGTCAAGGCCCTTGACCGCCATCTGGCAGGA 791
DB 75 GGTGAAGCAGTGGCCCTGACTGAGGAGAAAGTCAAGGCCCTGGTGAAGTCTGCACTGA 134

QY 792 GATGAGAGAGGAGGCAAGATCACAAGATCGGCCCGGAGAACCCCTACAACACCCCGT 851
DB 135 GATGAGAGAGGAGGCAAAATCTCAAGATTGGCCCGGAGAACCCCTACAACACCCCTGT 194

QY 852 GTTCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGAATTCGCGGAGCT 911
DB 195 GTTTGCCATCAAGAAGAGGACTCCACCAAGTGGAGGAAGCTGGTGAATTCAGGGAGCT 254

QY 912 GAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCGCCACCCCGCGGCT 971
DB 255 GAACAAGAGGACCCAGGACTTCTGGAGGTGAGCTGGGCATCCGCCACCCCGCGGCT 314

QY 972 GAAGAAGAGAGAGCGTGCACCGTGTGACGTGGCGCAGCGCTACTTCAGCGTGCCT 1031
DB 315 GAAGAAGAGAGAGTCTGTGACTGTGCTGGATGGGGGATGCTACTTCTGTGCCCCCT 374

QY 1032 GGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCTCAACAACAGAGACCCC 1091
DB 375 GGATGAGGACTTCAGGAAGTACACTGCCTTCCACCATCCCTCCATCAACAAATGAGACCCC 434

QY 1092 CGGCATCCGCTACAGTACAACTGTGCTGCCAGGCTGGAGGGGAGGCCCGCAGCATCTT 1151
DB 435 TGGCATCAGTACAGTACAACTGTGCTGCCAGGCTGGAGGGGATCCCTGCCATCTT 494

QY 1152 CCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCCGCGCCCGCAACCCCGAGATCGTGAT 1211
DB 495 CCAGTCTCCATGACCAAGATCTTGAGGCCCTTCCAGGAAGCAGAACCTGACATTTGTAT 554

QY 1212 CTACCA-----GGCCCCCTGTATCGTGGGAGCGACCTGGAGATCGGCAGCACCGCGC 1265
DB 555 CTACCAAGTACATGGATGACCTGTATGTGGGCTCTGACCTGGAGATTGGGAGCACAGGAC 614

```

```

QY 1266 CAAGATCGAGGAGCTGCGCAAGCAGCACTGCTGGCTGGGGCTTCACCAACCCCGACAAGAA 1325
DB 615 CAAGATTGAGGAGCTGAGGCAGACCTCTGCTGAGGTGGGGCTTGACCAACCCCTGACAAGAA 674

QY 1326 GCACCAGAAGAGAGCCCGCTTCTGCTCCCAT-----CGAGCTGCAACCCCGACAAGTGAAC 1379
DB 675 GCACAGAAGAGAGCCCGCTTCTGCTGATGGGCTATGAGCTGCAACCCCGACAAGTGAAC 734

QY 1380 CGTCAGCCCCATCGAGCTGCCGGAAGAGAGCTGGAACCGTGAACAGCATCCAGAAGCT 1439
DB 735 TGTGAGCCCCATTTGTGCTGCTGAGAGGACTCTCTGGAATGGAATGACATCCAGAAGCT 794

QY 1440 GGTGGGCAAGCTGAATGGGCGCAGCCAGATCTACCCCGGCATCAAGGTGAGGAGCTGTG 1499
DB 795 GGTGGGCAAGCTGAATGGGCGCTCCCAATCTACCCCTGGCATCAAGGTGAGGAGCTGTG 854

QY 1500 CAAAGCTGCTGGCGGCGCAAGGGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCGCA 1559
DB 855 CAAAGCTGCTGAGGGGCGACCAAGGGCCCTGACTGAGGTGATCCCTCTGACTGAGGAGGCTGA 914

QY 1560 GCTGAGCTGCGCGAGAAACCGCGAGATCCTGCGGAGCCCGTGCACGGCGTGTATACGA 1619
DB 915 GCTGAGCTGCTGAGAAACAGGGAGATCTGAAAGGAGCTGTGCAATGGGGTGTACTATGA 974

QY 1620 CCCCAGCAAGGACCTGGTGGCGGAGATCCAGAGCAGGGCCACGACCACTGAGTGGACCTACA 1679
DB 975 CCCCCTCCAAAGGACCTGATTGCTGAGATCCAGAAGCAGGGCCAGGGCCAGTGGACCTTACA 1034

QY 1680 GATCTACCAAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATCGCACCCG 1739
DB 1035 AATCTACCAAGAGCCCTTCAAGAACCTGAGACTGGAAGTATGCGAGATGAGGGGGG 1094

QY 1740 CCACACCAACGAGTGAAGAGCTGACCGAGGCGGTGAGAGATGCGCATGAGAGAGCAT 1799
DB 1095 CCACACCAATGATGTGAAGCAGCTGACTGAGGCTGTGCAAGAGATCAACCACTGAGTCCAT 1154

QY 1800 GGTGATCTGGGGCAGACCCCGAGTTCGGCTGCGCATCCAGAGGAGAGCTGGGAGAC 1859
DB 1155 TGTGATCTGGGGCAGACCCCGAGTTCGGCTGCGCATCCAGAGAGGAGACCTGGGAGAC 1214

QY 1860 CTGTGTGACCGACTACTGGCAGGCGCACCTGATCCCGAGTGGGAGTTCGTGAACACCC 1919
DB 1215 CTGTGTGACTGAGTACTGGCAGGCGCACCTGATCCCTGAGTGGGAGTTCGTGAACACCC 1274

QY 1920 CCCCCTGTGTAAGCTGTGGTACCACTGGAAGAGAGCCCATCATCGGGCGCGAGACCTT 1979
DB 1275 CCCCCTGTGTAAGCTGTGGTACCACTGGAAGAGAGGAGCCCATTTGTGGGGCTGAGACCTT 1334

QY 1980 CTACGTGACCGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGCGCTAGCTGACCGA 2039
DB 1335 CTATGTGATGGGGCTGCGCAACAGGGAGACCAAGCTGGGCGAAGGCTGGCTATGTGACCA 1394

QY 2040 CCGGGCGCGGAGAGATCGTGAGCTTGACCGAGACCAACCAACAGAGAGACCGAGCTGCA 2099
DB 1395 CAGGGCGAGGAGAGAGTGTGACCTGACTGACACCAACCAACAGAGAGACTGAGCTCCA 1454

QY 2100 GGCCATTCAGCTGGCCCTGCGAGGACAGCGGAGAGAGTGAACATCGTGAACCGACCGCA 2159
DB 1455 GGCCATTCAGCTGGCCCTCCAGGACTCTGGGCTGGAGGTGAACATTTGACTGACTGCCA 1514

QY 2160 GTACGCGCTGGGCATCATCCAGGGCCCGCGACAGAGAGCGAGCTGGTGAACCA 2219
DB 1515 GTATGCTGAGGAGTATCCAGGGCCCGCGCTGATCAGTCTGAGTCTGAGCTGGTGAACCA 1574

QY 2220 GATCATTCAGAGAGCTGATCAAGAGAGAGTGTACCTGAGCTGGGTGCGGCCCGACAA 2279
DB 1575 GATCATTCAGAGAGCTGATCAAGAGAGAGTGTGTACCTGAGCTGGGTGCGGCCCGACAA 1634

QY 2280 GGGCATTCGGGCGCAACGAGAGAGATCGAACAGTGTGTGAGCAAGGCGATCCGCAAGGTGT 2339
DB 1635 GGGCATTCGGGCGCAATGAGCAGGTGGAACAAGCTGTGTCTGCTGGCATCAGGAAGGTGTCT 1694

```

Qy 2340 GTTCCTGGACGGCATCGA 2357  
|||||||  
Db 1695 GTTCCTGGATGGCATTGA 1712

Search completed: June 2, 2005, 03:13:16  
Job time : 399.052 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 07:40:05 ; Search time 7836.86 Seconds  
(without alignments)  
11933.846 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457

Sequence: 1 gtcagccaccatggcgca.....gggtagcaccggggaattc 2457

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: gb\_est1.\*

2: gb\_est2.\*

3: gb\_hc.\*

4: gb\_est3.\*

5: gb\_est4.\*

6: gb\_est5.\*

7: gb\_est6.\*

8: gb\_gss1.\*

9: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	136.4	5.6	330	9	CL293849
2	100	4.1	2886	9	CL967755
3	90.2	3.7	2598	3	AY103647
4	88.8	3.6	3069	9	CL973991
5	86.6	3.5	1398	9	CL961989
6	85.6	3.5	951	4	BM321451
7	84.2	3.4	1941	9	CL971508
8	80.6	3.3	869	7	CK159167
9	80.2	3.3	1132	4	BM320864
10	79.6	3.2	1165	4	BM320900
11	79.6	3.2	1680	9	CL982770
12	79	3.2	867	4	BM321430
13	78.4	3.2	1725	9	CL978463
14	78.4	3.2	2031	9	CL974989
15	78.2	3.2	2697	9	CL952258
16	78.2	3.2	2853	9	CL974397
17	78	3.2	1485	9	CL970981
18	77.4	3.2	1509	9	CL959255
19	77.4	3.2	2559	9	CL982027
20	77	3.1	11691	9	CL962901
21	76.8	3.1	2682	9	CL969033
22	76.6	3.1	1550	4	BM321022
23	76.2	3.1	1401	9	CL962721
24	76.2	3.1	3249	9	CL945510

25	75	3.1	545	4	BI724851	BI724851	1031075E0
26	75	3.1	862	4	BM321023	rockefell	
27	74.4	3.0	914	8	BZ568300	pac92-164	
28	74.2	3.0	2313	9	CL982362	OsIFCC047	
29	74	3.0	1290	9	CL972679	OsIFCC023	
30	74	3.0	2072	3	CR603312	full-leng	
31	73.8	3.0	2433	9	AY401196	AY401196 Homo sapi	
32	73.2	3.0	853	4	BM321393	rockefell	
33	72.8	3.0	566	4	BM587428	BM587428 170006873	
34	72.6	3.0	788	6	CB643171	CB643171 OSJNEB03L	
35	72.6	3.0	849	3	CR718981	CR718981 Tetraodon	
36	72.4	2.9	753	3	CC675888	CC675888 OGWCO51TH	
37	72.4	2.9	2523	9	CL974879	CL974879 OsIFCC026	
38	72.2	2.9	602	7	CV057146	BNEL2463	
39	72.2	2.9	640	2	BE601575	BE601575 HVSMEH009	
40	72.2	2.9	688	6	CB648640	CB648640 OSJNEB12C	
41	72.2	2.9	764	6	CB651670	CB651670 OSJNEB16L	
42	72.2	2.9	766	6	CB642928	CB642928 OSJNEB03F	
43	72.2	2.9	809	6	CB641397	CB641397 OSJNEB01A	
44	72.2	2.9	841	6	CB651502	CB651502 OSJNEB16H	
45	72.2	2.9	851	9	CG260623	CG260623 OGWGE95TV	

#### ALIGNMENTS

RESULT 1  
CL293849/c  
LOCUS: CL293849 330 bp DNA linear GSS 12-FEB-2004  
DEFINITION: 02S0349-08A1-C03 UniformMu MutAIL Library Zea mays genomic clone  
ACCESSION: 02S0349-08A1-C03, genomic survey sequence.  
VERSION: CL293849  
KEYWORDS: GSS.  
SOURCE: Zea mays  
ORGANISM: Zea mays  
REFERENCE: 1 (bases 1 to 330)  
AUTHORS: Latshaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.  
TITLE: Sequence tagged transposon insertions from the UniformMu maize population  
JOURNAL: Unpublished (2003)  
COMMENT: Contact: Donald R. McCarty  
Plant Molecular and Cellular Biology Program  
University of Florida  
PO 110690 Gainesville, FL 32611-0690, USA  
Tel: 352-392-1928 x322  
Email: drmc@ufl.edu

FEATURES  
source  
1. .330  
Sequence flanking probable Mu insertion site in UniformMu line:  
02S0349-08, Primer set: A  
Class: transposon insertion site.  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="W22 (ACR, bz1-m9)"  
/cultivar="UniformMu"  
/db\_xref="taxon:4577"  
/clone="02S0349-08A1-C03"  
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."

#### ORIGIN

Query Match 5.6%; Score 136.4; DB 9; Length 330;  
Best Local Similarity 66.4%; Pred. No. 2.5e-14;

Matches 211; Conservative 0; Mismatches 106; Indels 1; Gaps 1;  
 QY 421 GCGGCCAGATCAAGAGCCCTGCTGGACACCGGCGGAGCACACCGTGTGGAGAG 480  
 Db |||||  
 Db 329 GGGGGGAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA 270  
 QY 481 ATGAGCTCCCGCGCAAGTGAAGCCCAAGATGATCGGGCGCATCGGGGCTTCATCAAG 540  
 Db |||||  
 Db 269 ATGACTTTGACAGGAAGATGGAACCAAAATGATAGGGGAATGGAGGTTTATCAAA 210  
 QY 541 GTGCGCCAGTACGACCATCTGATCGAGATCTGCGGCAAGAGCCATCGGCACCGTG 600  
 Db |||||  
 Db 209 GTAAGACAGTATGATCAGTACCCATAGAAATCTGTGGGCAATAGCTATAGTACGGTA 150  
 QY 601 CTGATCGGCCCCACCCCGTGAAATCATCTGCGCGGCAATCTGACCCAGCTGGGCTGC 660  
 Db |||||  
 Db 149 TTATAGGACCTACACCTGTCAACATAATTGGAAGAAATCTTTGACTCAGATTGG-TGC 91  
 QY 661 ACCCTGAACTTCCCATAGCCCATCGAGACCGTGCCGCTGAAGCTGAAGCCCGGCATG 720  
 Db |||||  
 Db 90 ACCTTAAATTTTCCCATAGTCTTATTTGAACTGTACAGTAAATTAAGCCAGGAATG 31  
 QY 721 GACGGCCCCAAGGTGAAG 738  
 Db |||||  
 Db 30 GATGCCCCAAGTAAGG 13

RESULT 2  
 CL967755  
 LOCUS  
 DEFINITION  
 OaIFCC015718 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Oryza sativa (indica cultivar-group)  
 Oryza sativa (indica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.  
 REFERENCE  
 AUTHORS  
 Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L., Wong,G.K.S., Deng,X.W. and Wang,J.  
 An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
 JOURNAL  
 COMMENT  
 Unpublished (2004)  
 Contact: Chen Chen  
 Department of Bioinformatic  
 Beijing Institute of Genomics  
 Chinese Academy of Sciences, Beijing 101300, China  
 Tel: 86-10-80481559  
 Fax: 86-10-80488676  
 Email: chenchen@genomics.org.cn  
 Rice genomic sequence.  
 Class: exon-trapped.  
 FEATURES  
 source  
 1..2886  
 /organism="Oryza sativa (indica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:39946"  
 /clone\_lib="Oryza sativa Express Library"  
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN  
 Query Match 4.1%; Score 100; DB 9; Length 2886;  
 Best Local Similarity 42.8%; Pred. No. 9.6e-08;  
 Matches 700; Conservative 0; Mismatches 915; Indels 21; Gaps 3;  
 QY 189 GATGAAGAGCTGACCGAGCGCAGCCAACTTCTTCGCGAGGACCTGGCCCTTCCCCCA 248  
 Db |||||  
 Db 219 GGTCTAGGAGCGCGCCCATCATGGCCATCTCTCGCCAAACCGCGGGCAGGCGCGCC 278

QY 249 GGGCAAGGCCCGCGAGTTCCCCAGCGAGCAGAACCGCGCCCAACAGCCCAACAGCGCGCA 308  
 Db |||||  
 Db 279 CGACTGGGAGAGACTTCGTGGCATCGTCTCTCTCTCATCACTCAACCATCAGTTT 338  
 QY 309 GCTCAGGTGGGC---GGCGACAACCCCGCAGCGAGCGCGCGCGCGCGCAGCGCAC 365  
 Db |||||  
 Db 339 CATCGAGGAGAACATGCGCGGAAACCGCGCGCTCATGGCGCGCTCGCGCCCA 398  
 QY 366 CCTGAATTCCTCCCGATACCTGTGCGAGGCCCCCTGTGTGAGCATCAAGTGGGCGG 425  
 Db |||||  
 Db 399 GACCAAGTGTCTAGGGACGGGAAATGGCAGGAGCAGGACGCTCATCTCTGTTCCCGG 458  
 QY 426 CCAGATCAAGGAGGCCCTGTGGACACCGCGCGCGACACACCGTGTGGAGGAGATGAG 485  
 Db |||||  
 Db 459 CGACATCATAGCATCAAGCTCGGGGACATCATCCCCCGCGACGCGCGCTGTCTCGAGG 518  
 QY 486 CTGCCCCGCAAGTGGAAAGCCCAAGATGATCGGGGACATCGGGCGCTTTCATCAAGGTGG 545  
 Db |||||  
 Db 519 CGACCGCTCAAGGTTCGACCGGCGGCTCACCGCGAGTCGATGCCCGCTCAACAAGCA 578  
 QY 546 CGAGTACGACAGATCTGTGATCGAGATCTGCGGCGAAGAGGCATCGGCACCGTCTGAT 605  
 Db |||||  
 Db 579 CGCGGGCAGGGCGCTTCTCCGGTCCACCGTCAAGCAGGGCGAGATCGAGGCGCTCGT 638  
 QY 606 CGGCCCCACCCCGTGAACATCATCGGCGCGCAACATCTGACCCAGCTGGGCTGCACCT 665  
 Db |||||  
 Db 639 CATGCCACCGCGGTGCACCTTCTTCGGAAGCGCGGCGCACCTGTGTGGAAGCAGCCNA 698  
 QY 666 GAATTTCCCATCAGCCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATGAGCGG 725  
 Db |||||  
 Db 699 CAACATCGGCACCTTCAGAGCTCGTCTCACGGCCATCGGCAACTTCTGCATCATCTCCAT 758  
 QY 726 CCCCAAGGTGAAGTCGTGCGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTG 785  
 Db |||||  
 Db 759 CGGCGTGGCATGATAATCGAGATCATCTTATATACCCGATCCAGCACCGCGCTACCG 818  
 QY 786 CGAGGAGATGAGAGGAGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTACAACAC 845  
 Db |||||  
 Db 819 CGACGGCATCGACAACCTGCTGCTCTCTCTATCGGCGGCGATCCCATCGCCATGCCAC 878  
 QY 846 CCCCGTGTCCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCG 905  
 Db |||||  
 Db 879 CGTGTGCTCGTACCATGGCCATCGGCTCCACCGGCTGTCCGACGAGGCGGCCATCAC 938  
 QY 906 CGAGCTGAACAGCGCACCGAGACTTCTGGAGGTGACAGCTGGGCATCCCCCAACCCGCG 965  
 Db |||||  
 Db 939 CAAGCGATGACCGCCATCGAGGAGATGCGCGGCATGGACGCTCTCGACGCAAGAAGAC 998  
 QY 966 CGGCTGAAGAAGAGAGCGTGACCGTGTGGACGTGGCGGACGCTACTTCAGCGT 1025  
 Db |||||  
 Db 999 CGGAGCTGACGCTCAACAAGCTCACCGTCGACAGAGAGCTGATCGAGGTGTACGGGG 1058  
 QY 1026 GCCCTTGGAGGAGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAACGA 1085  
 Db |||||  
 Db 1059 GGGGCTGGACAAGGACTCGGTGCTCTGTATCGCGCGAGGGCGTCCCGCTCGAGAACA 1118  
 QY 1086 GACCCCCGGCATCCGCTACAGATCAACGCTGCTGCCCGAGGGCTGGAAGGGAGCCCCAG 1145  
 Db |||||  
 Db 1119 GGAGCCCATTTGACACATCATCTGGGGCATGCTCGCCGACCCCAAGAGGGCCCCGCGCGG 1178  
 QY 1146 CATCTTCAGAGCAGCATGACCAAGATCTTGAGGCGCTTCGCGCGCCGCAACCCCGAGAT 1205  
 Db |||||  
 Db 1179 CATGAAGAGGTCTCACTTCCTC-----CCGTTCAACCCCGTTCGAGAAGGCGAC 1226  
 QY 1206 CGTGATCTTACAGGCCCCCTGTAGTGGGCGAGCGACCTGGAGATCGGCGAGCACCGCGC 1265  
 Db |||||  
 Db 1227 GSCCATCACTTACATCGACGGCAATGGGAATGGACAGGATCAGAAAGGGCGCGCGGA 1286  
 QY 1266 CAAGATCGAGAGCTGCGCAAGACCTCTGCTGGCTGGGGCTTCACCAACCCCGCAAGAA 1325  
 Db |||||  
 Db 1287 GCAGATCATGAGCTGTGCAAGATGAGCAAGGAGCGCCGAGAAAGGTGACACCTGAT 1346  
 QY 1326 GCACCAAGAGGAGCCCCCTTCTTCCCATCGAGCTGCACCCCGACAAAGTGGACCGTGCA 1385

```
Db 1347 CGACACGATCGCGGACCGCGGCTCCGGTGGCTGGGCTGTGTACCAAGAGGTGCCGGA 1406
Qy 1386 GCCATCGAGCTGCCGAGAGAGAGAGAGTGGACCGTGAACGACATCCAGAAAGCTGGTGGG 1445
Db 1407 GAAGAGCAAGAGAGAGAGAGAGAGGAGCCATGCGAGTTCTGTGGCTCTCTGCGCTGTTCGA 1466
Qy 1446 CAAGCTGAATGGGCGACGACAGATCTACCCGCGCATCAAGGTGGCGCCAGCTGTGCAAGCT 1505
Db 1467 CCGCGGAGGACGACAGCGCGGAGACCATCGCGCGCGCTGCACTTCGGCGTGAACGT 1526
Qy 1506 GCTCGCGCGCGCAAGGCGCTGACCGACATCGTGCCTGACCGAGGAGGCGGAGCTGGA 1565
Db 1527 GAAGATGATACCGCGGACGAGCTCGCCATCG-----GGAAGAGAGCGGCGCGCGCT 1580
Qy 1566 GCTCGCGGAGAACCGCGAGATCTCTGCGGAGCGCGTGCACGCGCGTGTACTACGACCCCGAG 1625
Db 1591 CGGGATGGGACCAACATGTACCCGTGACGACGCTGTGTGGCGCAAGAGAGCGGAGAT 1640
Qy 1626 CAAGACCTGTGGCGGAGATCCAGAACGAGGAGGCGGACGAGTGGAGCTTACCAAGATCTA 1685
Db 1641 GAGCGGCTCCCAATCGAGAGCTGATCGAAGAGGCGGAGCGCTTGGCGCGCTGTTCCT 1700
Qy 1686 CCAGGAGCCCTTCAGAGACCTGAAGACCGGCAAGTACCGAAGATGCCAGTGCACGCGCCAC 1745
Db 1701 GGAGCAAGTACGAGATCGTGAAGCGCTCCAGGACCGGAAGCAGCATCTCGCGCATGAC 1760
Qy 1746 CAACGACGTGAAGCAGCTGACCGGCGGTGCAGAAAGATGCCATGGAGAGCATCTGTGAT 1805
Db 1761 CGCGAGCGGCTGAACGAGCGCGCGCGCTGAAGAGGCGGCTGAAGAGGCGGACATCGCGCTG 1820
Qy 1806 CTGGGCGCAAGACCCCT 1821
Db 1821 CGACGCCACGAGCGCC 1836
```

## RESULT 3

```
AY103647
LOCUS AY103647 2598 bp mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PC0142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2598)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 2598)
Coe,E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
Location/Qualifiers
1. 2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus
```

## FEATURES

source

## Library"

/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed DuPont contigs; this resource was  
assembled by DuPont as part of a collaboration for the  
overgo addressing of BACs in conjunction with the Maize  
Mapping Project."

## ORIGIN

```
Query Match 3.7%; Score 90.2; DB 3; Length 2598;
Best Local Similarity 42.2%; Pred. No. 5, 4e-06;
Matches 814; Conservative 0; Mismatches 1098; Indels 15; Gaps 5;
Qy 296 CCACAGCGCGAGCTGCGAGTGGCGGCGCAACCCCGCAGGAGCGCGCGCGAGC 355
Db 51 CCACACCTCTCCCTCTGCGCACCGCGCGGCCCAACCAACACACCGAGCGAGCAA 110
Qy 356 GCAGGGGACCTGAACTTCCCCAGATCACCTGTGTGCGAGCGCCCCCTGTGTAGCATCA 415
Db 111 TGGCGGGCAACCGGCGCCATCGTGGAGAGCGACCCGCTGAATGGGGCGCGCGCGGG 170
Qy 416 AGTGGGCGGCGCAGATCAAGGAGGCCCTGTCTGGACACCGCGCGCGACACACCGTGTGG 475
Db 171 AGCTGGCGGGAGCCACTTGAACGAGGTGAAGCGCATGTGTGCGAGGCCCGCGAGCCGG 230
Qy 476 AGGAGATGAGCTGCCCGGCAAGTGAAGGCCCAAGATGATCGGGGCGCATCGCGCGCTTCA 535
Db 231 TGGTCAAGATCGAGGGTTCCACCCCTCGCGTGGCGGAGTGGCGCGCTCGCTCCGCGCA 290
Qy 536 TCAAGTGGCGCAGTACGACGACGATCTGTGATCGAGATCTGGGCGAAGAGCGCATCGGCA 595
Db 291 AGGACGCGTCCGGCGTCCCGCTCGAGCTCGACGAGGAGCGCGCGCGCGCTCAAGGCGCA 350
Qy 596 CCGTGTGATCGGCGCCACACCCCGTGAACATCATCTGGCGCGCAACATGCTGACCCAGCTGG 655
Db 351 GCACGAGTGGATCTCTGATGTCATCGCCACGGGCGGCACTTACCGCGTACCACGG 410
Qy 656 GCTGCACCTGTAACCTTCCCATCGACCCCATCGAGACCGTGTCCGTGAAGCTGAAGCCCG 715
Db 411 GCTTGGCGGCGACCTCCACCGCGCAACCAAGGAGCGGCGCGCTCGAGTTCGAGCTGC 470
Qy 716 GCATGGAGGCGCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCCCTGA 775
Db 471 TCAGGSCATCTCAACCGCGGAATCTTCGGCAACCGGACGCGGCGCACACGCTGCGTGG 530
Qy 776 CCGCATCTCGAGAGAGATGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 835
Db 531 AGGTACCGCGCGCGCGATGCTGTGCGCATCAACACCCCTCTTCAGGGGCTACTCCGCGCA 590
Qy 836 CCTACAACACCCCGCTGTTCGCCATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 895
Db 591 TCCGCTTCGAGATCTCGAGGCCATCAAGAGCTGCTCAACACCGGTGTTCAGCCCTGCC 650
Qy 896 TGGACTTCGCGGAGCTGAACAGGCGACCCAGGACTTCTGGGAGAGTGTGAGCTGGGCGATCC 955
Db 651 TGGCGCTCCGGGCGACCATCACCGCTCGGCGGAGCTGTGTCCTCTTACATCGCGCG 710
Qy 956 CCACCCCGCGCGCTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1015
Db 711 GCCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 767
Qy 1016 ACTTCAGCGTGCCTCGAGCGAGGACTTTCGCAAGTATACACCGCTTACCATCCCGAGCA 1075
Db 768 ACGCGCGCGAGGCGTTCAGATCGCGCGCATCGAGGGCGGCTTCTTCAAGCTCAACCCCA 827
Qy 1076 TCACAAACGAGAGACCCCGGCGATTCGCTACAGTACAGCTGTGCGCGCGCGGCTGGAAGG 1135
Db 828 AGGAGGCGCTCGCATCGTCAACCGGCGAGTCCGCTGGGCTCGCGCTCGCGCGCGCGCTGA 887
Qy 1136 GCAGCCCGAGCATCTTCAGAGGAGCATGACCAAGATCTCGAGCCCTTCGCGCGCGCGCA 1195
Db 888 TGTACGACGCGCAACGCTCTCGCGCGCTGTGTGCGAGGCTCT---GTCCCGCGCTTCTGCG 944
```

Qy	1196	ACCCGAGATCGTGATCTA	CGAGCCCCCCTGTACTGTGGCAGCGACTCTGGAGATCGGCC	1255
Db	945	AGGTCATGAACGGCAAGCCCGAGTACACGACACACCTGACCCACAAGCTGAAGCACCACC	1004	
Qy	1256	AGCACCGGCCAAGATCGAGGAGTCGCGAAGCAACCTGCTG---	CGTGGGGCTTACCA	1312
Db	1005	CGGGTCCATCTGAGGCCCGGCCCATCATGAGCACAATCTCTGGATGGCAGCTCTTTCATGA	1064	
Qy	1313	CCCCGACAAGAAGCACAGAAGGAGCCCCCTTCTGTCGCCATCTGAGCTGCACCCCGACA	1372	
Db	1065	AGCAGGCCAAGAGTTGAACGAGCTGGAACCGCTGCTGAAGCCCAAGCAGCAGCAGTATCG	1124	
Qy	1373	AGTGGACCGTCAGCCCCATCGAGCTGCCGAGAGGAGAGCTGAGCCGTGAACGACATCC	1432	
Db	1125	CGCTCCGACGTCGCGCAGTGGCTGGGCCCCCAAGATCGAGGTCATCCGCGCGCCACCA	1184	
Qy	1433	AGAAGCTGGTGGGCAAGCTGAACCTGGCCAGCCAGATCTACCCGGGATCAAGGTGGGCC	1492	
Db	1185	AGTCCATCGAGCGCGAGGTCAACTCCGTGAACGA---	CAACCCCGTTCATCGAGCTCCACC	1241
Qy	1493	AGCTGTCAAGCTGCTCGCGGCCCAAGGCCCTGACCGCATCTGTGCCCTTCACCGAGG	1552	
Db	1242	CGGCAAGGCGCTGCACGGCGGCACATTCAGGGCACCCCATCGGGGTCTCATGGACA	1301	
Qy	1553	AGGCCGAGCTGGAGCTGGCGAGAACCGCGAGATCTTGCAGCGCCCGTGCACGGGGTGT	1612	
Db	1302	ACGCCCGCTCGCCATCGCCAAATCGGCAAGCTCATGTTCCGCGAGTTCTCCGAGCTCG	1361	
Qy	1613	ACTACGACCCGAGCAAGGACCTGTGTGCGAGATCCAGAAGCAGGGCCACGACAGTGGGA	1672	
Db	1362	TCAACGAGTTCTACAACAAGGGGTCACTTCAACCTTGGCGGCGAGCGCAACCCCGACC	1421	
Qy	1673	CCCTACAGATCTACCGAGGCCCTTCAAGAACCTGAGACGGGCAAGTAGCGCCAAGATGC	1732	
Db	1422	TGGACTACGGCTTCAAGGGCACCGAGATCGCATGGCTCTTACTGTCTCGAGCTCCAGT	1481	
Qy	1733	GCACCGGCCACACCAACGAGCGTGAAGCAGCTGACCGAGGCCGTGCAGAGATCGCCATGG	1792	
Db	1482	ACTGGGCAACCCATCACCAAACCATGTCAGAGCGCGGAGCAGCACACCAGGACGTGA	1541	
Qy	1793	AGAGCAT---CGTGATCTGGGGCAAGACCCCAAGTTTCGCGCTGCCCATCCAGAAAGAGA	1849	
Db	1542	ACTCCCTGGGCCTCGTCTCGGCCAGGAAGACCGCGGAGGCGATCGACATCTCTGAAGCTCA	1601	
Qy	1850	CCTGGGAGACTGTGTGACCGACTACTTGGCAGGCCACCTGGATCCCAGATGGGAGTTCG	1909	
Db	1602	TGTCGTCCACTACATCGTGGCGCTGTGCCAGGCCGTGGACCTGCGCCACTCGAGAGA	1661	
Qy	1910	TGAACACCCCCCTCGTGAAGCTGTGTGTACACAGCTGGAGAGGAGCCCATCATCGCGC	1969	
Db	1662	ACATCAAGCGCTCGGTGAAGAACACCGTGACCCAGGTGGCCAGAGAGGTCTGACCATGA	1721	
Qy	1970	CCGAGACTTCTACGTGGAAGCGGCCGCCAAACCGCGAGACCAAGATCGGCAAGCCGGCT	2029	
Db	1722	ACCCTCGGGGAGCTCTCCAGCGCCCGCTTCAGCGAAGAGGAGCTGATCAGCGCCATCG	1781	
Qy	2030	ACGTGACGACCGGGCCGCGAGAGATCTGAGCTTGAACGAGACACCAACGAGAGA	2089	
Db	1782	ACCOCAGGCGGTTCACGTACGCGGAGGACGCGGCCAGCGGAGCTTCCCTCGATGC	1841	
Qy	2090	CCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGCAGCGAGGTGAACATCGTGA	2149	
Db	1842	AGAAGCTGCGGCCGCTGCTGTGTGAACCAACGCCCTCAGACGGCGCACCGGAGCGGGAGC	1901	
Qy	2150	CCGACGACGATAGCCCTTGGGCATCATCCAGGCCACCGCCGACAAAGAGCGAGCGAGC	2209	
Db	1902	CCTCGGTGTTCTCCAAGATCACAGGTTTCGAGGAGGAGCTCCGCGCGTCTGCCCCAGG	1961	
Qy	2210	TGGTGAA	2216	
Db	1962	AGTGGGA	1968	

RESULT 4	CL973991	3069 bp	DNA	linear	GSS 21-SEP-2004
LOCUS	CL973991				
DEFINITION	OsIFC025416 <i>Oryza sativa</i> Express Library <i>Oryza sativa</i> (indica cultivar-group) genomic, genomic survey sequence.				
ACCESSION	CL973991				
VERSION	CL973991.1	GI:52402507			
KEYWORDS	GSS.				
SOURCE	<i>Oryza sativa</i> (indica cultivar-group)				
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; <i>Oryza</i> .				
AUTHORS	1 (bases 1 to 3069) Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G.K.S., Deng, X.W. and Wang, J.				
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis				
COMMENT	Unpublished (2004) Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped. Location/Qualifiers 1..3069 /organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_lib="Oryza sativa Express Library" /note="Oryza sativa exon trapped genomic sequences "				
ORIGIN					
	Query Match	3.6%	Score 88.8	DB 9	Length 3069;
	Best Local Similarity	41.5%	Pred. No. 9.8e-06;		
	Matches 652;	Conservative 0;	Mismatches 917;	Indels 3;	Gaps 1;
Qy	37	ACCAGCGCCAACATCTCTGATGCAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAG	96		
Db	625	AGCGGCATCGCGAAGACGACGCTGGCGAGCTACTCTACCGAGGACACGATCCGGAAG	684		
Qy	97	TGCTTCAACTGCGGCAAGAGGGGCACATCGCCGCGCACTGCGCGCCCCCGCGAAGAAG	156		
Db	685	CAC TTCGCTTCGACGGCGTGCGGTCCACGTCCCGCGCAAGTTCGCGTTCGCGGACCTCTC	744		
Qy	157	GGCTGTGTGAAGTGCAGGAGGGGCCACAGATGAAGACTGCACCGAGCGCCAGGCC	216		
Db	745	CACAGCTCATTCGCGCAGGCCAAGCCTCTCTCGGCGACGACGAGCTCGCGGGGAGGC-	803		
Qy	217	AAC TTCTTCGCGAGGACCTTGGCCTTCTCCCGCAGGGCAAGCCCGCGAGTTTCCCCAGCGAG	276		
Db	804	--CGCCACGACGACGTCGCGCGCGCGGGTGGATGAGCGGAGAGGCTCGTCAAGCGC	861		
Qy	277	CAGAACCGCGGCACAGCCCCACAGCCGCGAGCTGCAAGTGGCGGCGACAAACCCCGC	336		
Db	862	CGCCTCGCCCGCGCTTCACAGGGGGAAGGATCCTCGTGGTGTCTCGACGACGCTCCGGAGC	921		
Qy	337	AGCAGGCGCGCGCAGCGCGCAGGGGCACCTGAAC TTCCCGCAGATCACCTGTGCGAG	396		
Db	922	AGGAGAGAGTGGGCGCTTCTTCTCGCGCGCTCCCGCGCGGCTTCGCGGAGCTGGCTG	981		
Qy	397	CGCCCCCTGTGTAGCATCAAGAGTGGCGGCGCAGATCAAGAGAGGCCCTTCTGTGACACCGCGC	456		
Db	982	CTCGTCACGACGAGGTCAGATCAGCGAGGATCGGAGCAACCGGTGGCGCGCGCGC	1041		
Qy	457	GC CGACGACACCGTGCTGTGAGGAGATGAGCCTGCCCGGCAAGTGGAAGCCAGATGATC	516		
Db	1042	GGCCTCTGTGGCACACGAGGAGCTCGGGAAGCTCGGCGCGGAGAGGCGCGGAGCTGTC	1101		

```
Qy 517 GCGGCGCATCGCGGCTTTCATCAAGGTGGCGCAGTACGACCAAGCTCTGATCGAGATCTGC 576
Db 1102 CGCGGAGGGTGTAGCGCCACCGGAGCGGAGCGGAGAGATCGAGACGCTCAAGAGC 1161
Qy 577 GCGAAGAGGCGATCGGACCGTGTGTATGGCCCGCCACCGCGTGAACATCATCGCGCGC 636
Db 1162 CTCGTGGACTCCATGACGAAGGGCTCCAACTGCGCTCAACATCGTCAATCTGCGCGGA 1221
Qy 637 AACATGCTGACCCAGCTGGGTGACCGCTGAACCTTCCCATCAGCCCATCGAGACCGTG 696
Db 1222 CTCCTGAGTCCAAGAGAGGATGATGGAGGCGGTGATCCATAGCTCGATGATACG 1281
Qy 697 CCCGTGAAGTGAAGCCCGCATGACGCGCCCGCCCAAGGTGAAGCAGTGGCCCTGACCGAG 756
Db 1282 CCACCCCGACATCGCGGAGCAGACGCGCGCGGAGCGGAGCGCGAAGAGCAGCCTC 1341
Qy 757 GAGAAGATCAAGCGCTGACCGCATCTGCGAGAGATGGAGAGAGGGGCAAGATAC 816
Db 1342 GATGATAGCGAGCGGCGCGGACATCGACGGAGCAACGAAGAAGACGAAGAGAGGAAG 1401
Qy 817 AAGATCGGCGCGAGAACCCCTACAAACACCCCGTGTTCGCCCATCAAGAAGAGCAGC 876
Db 1402 CAGATGGCGACATGATGGAACAAGATCTGACGGTGTGATGGACGACTCTCCGACGAT 1461
Qy 877 ACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAAGCGCACCCAGGACTTCTGG 936
Db 1462 CTGAAGCGTGTCTTCTTCTTCTTCCCGGGTTACCGCGCAGACCGCATCTGCGCGGG 1521
Qy 937 GAGTGCAGCTGGGGATCCCCACCGCGCGGCTGAAGAAAGAGAGAGCGTCAACCGTG 996
Db 1522 AAGCTGTGCTGGCTGTGGCTCGCGGAGGGTTCATGAGGCGAAGAACCGGCGACAGCGTG 1581
Qy 997 CTGCACTGGCGAGCGCTTACTTCAAGCTGCGCTTGGACGAGGACTTCCGCAAGTACAC 1056
Db 1582 GAGGAGCAGCGCGAGGAGTGCTCAAGAGCTCATCTCCCGGTGCTCTGCTCCAGCTCGTC 1641
Qy 1057 GCCTTCACTCCAGCATCAACAGAGACCCCGCATCTCGCTACCACTACAGTCAACAGTG 1116
Db 1642 GAGAGCAGCGCGGCGGAGGGTGCCTGCTGAGCATCCACAGCGGTCTGAC 1701
Qy 1117 CTGCCCCAGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTTG 1176
Db 1702 TTCTCCAGGCGGAGCGCGCAGACCACTTCTCCAGTGACAGCGGCGCGCGGC 1761
Qy 1177 GAGCCTTCCGCGCGCGCAACCCAGATCTGATCTACAGGCGCCCTGTAGTGGGC 1236
Db 1762 CTCTCCAAAGCGCGCGCGCGCTCGCGCTCCGCAACACTTACGACACCGACCTCGCC 1821
Qy 1237 AGCGACCTGGAGATCGGCGCAGCAGCGGCCAAGATCGAGGAGCTGCGCAAGCAGCTGTG 1296
Db 1822 GTCATGCTCGAGGCGCCAGCTCAACCTCTCTGCGCATCCCGAGCGCGCGCC 1881
Qy 1297 CGTGGGGCTTCAACACCCCGCAAGAAGCAGCAGAGGAGCCCGCTTCTGCCCATC 1356
Db 1882 GCCAGCGCGCGCGCGCGGTGGCAGAGAGCGCTCGAGCTCATCAATGGCGCGCC 1941
Qy 1357 GAGTGCACCCGACAAAGTGAACCGTGCAGCCCATGAGTGTCCCGAGAGAGGAGCTGG 1416
Db 1942 CCCACCTTCTCCGTCACCGGTTCGAGTTCCTCCGCGTGTGATCTCAAGGGCGGTGCGG 2001
Qy 1417 ACCGTGAACGACATCCAGAGCTGTGGCAAGCTGAACCTGGGCGCAGCGAGATCTACCCC 1476
Db 2002 CTGCGGACACGGGAGTGTCTGCGGAGAGATCGGTGGTGTATCTCACTCAGGTACTG 2061
Qy 1477 GGCATCAAGGTGCGCAGCTGTGAAGCTGTGCGCGCGCGCAGAGGCGCTTGACCGACATC 1536
Db 2062 GGCCTCTCGACACCAACATGAGGCACCTCCGAGCTCCGTCAAGAGGCTCCGGAACCTG 2121
Qy 1537 GTGCCCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGAGAAACCGCGAGATCTCTGCGGAG 1596
Db 2122 CAGACGTGAGCTGTGCGCGCAGCGAGCGTGTGAGCGCTGCGGTGCGGCTCTGCGCGAAC 2181
```

```
Qy 1597 CCCGTGCACGCG 1608
Db 2182 CCCTCGCTCGC 2193

RESULT 5
CL961989 1398 bp DNA linear GSS 21-SEP-2004
LOCUS OaIFCC006991 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
DEFINITION
ACCESSION CL961989
VERSION CL961989.1 GI:52378720
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 1398)
AUTHORS Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source
location/Qualifiers
1..1398
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN
Query Match 3.5%; Score 86.6; DB 9; Length 1398;
Best Local Similarity 45.9%; Pred. No. 2.3e-05;
Matches 508; Conservative 0; Mismatches 569; Indels 30; Gaps 5;

Qy 1 GTCGACGCCACCATGGCGGCGCATGAGCGAGCCACCGCAGCCACCGCCCAACATCTCTGATGCG 60
Db 166 GTCGAGAGGCGAGCTCGCGGAGCGGCGACATGCTTCTGCTCCGTGACACACCCACCAAG 225
Qy 61 CGCAGCACTTCAAGGCGCCCGACGCGCATCATCAAGTGTCTTAAGTCTTCAAGTCTGCGGAGGAGG 120
Db 226 GCGCGCGGCTCGGCGCGCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282
Qy 121 CACATCGCGCGCAACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 283 CGGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 342
Qy 181 GGCCACCAAGATGAAGGACTGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 343 GTCAGAGCCGCGGAGGCGCATCGAGAGATCTTCAGACCAACCGCGCGCGCGCGCGCGCGCG 402
Qy 241 TTCCCCCAGGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 403 TTCCAGATCTCTCTCAACCGCGGAGTTCCTCGCGGAGGCGCGCGCGCGCGCGCGCGCGCG 462
Qy 301 AGCGCGGAGTGCAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 463 GCGCCCGA-----CCGCGTCTTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
Qy 361 GGCACCTGAACTTCCCGCGAGATCACCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
```

Db 514 GCCCTCCAGCGCTCAAGGACGTGTACGCCGATGGTCCCGAGGAGGATC---CTC 570  
 Qy 421 GCGGCCAGATCAAGAGGCGCTGTGGACACGGGCGCGACGACACCGTGTGGAGGAG 480  
 Db 571 ACCACCAACTCTGTGTCGCGGAGCTGTCAAGCTCGCGCGCAACGCGTTCCTGGCCCCAG 630  
 Qy 481 ATGAGCTCTCCCGGCAAGTGGAAAGCCCAAGATGATCGGGCGCATCGGCGGCTTCATCAAG 540  
 Db 631 AGGATCTGTCTCGTCAACGCCATGTCCGCGCTCTGGAGGCCACCGGCGCGACGTGCGC 690  
 Qy 541 GTGCGCCAGTACGACGAGATCTGTATCGA---GATCTCGGCAAGAAGCCATCGGCACC 597  
 Db 691 GAGGTGCGCTACGCGTGGGCAAGACTCCAGGATCGGCGCCCAAGTTTCTCAACGCCAGC 750  
 Qy 598 GTGCTGATCGGCGCCACCCCGTGAACATCATCTGGCGCGCAACATGCTGACCCAGCTGGGC 657  
 Db 751 GTGGGATTCGGTGGTCTGCTTCAGAAAGACATCTCAACCTGCTG----- 798  
 Qy 658 TGCACCTGAACTTCCCCATCAGCCCCATCGAGACCGTSCCGTGAAGCTGAAGCCCGGC 717  
 Db 799 TACATCTGGAGTGAATGGCTCTCCCGAGGTGGCAACTACTTGAAGCAGGTATCAAG 858  
 Qy 718 ATGAGCGGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAGGAGAAGATCAAGGCCCTGACC 777  
 Db 859 ATCAAGCACTACCAAGAGAGCGGTTCGTGAACCGCGTGTCTGTCATGTTCAACACC 918  
 Qy 778 GCATCTGGAGAGATGGAAGAGAGGCGCAAGATCAACAGATCGGCCCCCGAGAACCC 837  
 Db 919 GTGGCGGCAAGAAGATCCCGTGTGGGCTTTCGCTTTCAGAAAGGACACACCGCGACACC 978  
 Qy 838 TACAACACCCCGTGTGGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTG 897  
 Db 979 AGGAGAGCGCCCGCATCAAGCTCTGCAAGGGGCTGATCGGCGCAAGGCCAAGGTGAGC 1038  
 Qy 898 GACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCGATCCCC 957  
 Db 1039 ATCTACGACCCCGAGTACGAGGAGGACCAAGTCCAGCGGACCTCGCCATGAGCAAGTTC 1098  
 Qy 958 CACCCGCGCGCTGAGNAGAGAGAGCGGTGACGCTGTGGAGTGGGCGACGCTTAC 1017  
 Db 1099 GATGGGATCACCCGCTCATCTCCAGCGGATGAGCGCCACCGCCATCAAGCAGGTGTC 1158  
 Qy 1018 TTCAGCGTCCCGCTCGACGAGACTTCGCAAGTACACCGCTTCACCATCCCGAGCATC 1077  
 Db 1159 GTGCGCTGGAGCGCTACAGAGCGGCGGAGGAGCGGCGGCTGTGATCTCACCAG 1218  
 Qy 1078 AACACGAGACCCCGGCTACCGCTAC 1104  
 Db 1219 TGGGACGAGTTCAGGAGCCTCGACTAC 1245

RESULT 6  
 BM321451  
 LOCUS  
 DEFINITION  
 rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library  
 Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase [EC  
 3.3.1.1], mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
 Mastigamoeba balamuthi  
 1 (bases 1 to 951)  
 Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and  
 Philippe,H.  
 The analysis of 100 genes supports the grouping of three highly  
 divergent amoebae: Dictyostellium, Entamoeba, and Mastigamoeba  
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
 11819461  
 11830664  
 CONTACT: Muller Miklos

Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockvax.rockefeller.edu  
 Insert Length: 951 Std Error: 0.00  
 POLYA=No.  
 FEATURES  
 Location/Qualifiers  
 1. 951  
 /organism="Mastigamoeba balamuthi"  
 /mol\_type="mRNA"  
 /strain="ATCC 30984"  
 /db\_xref="taxon:108607"  
 /clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
 /note="syn: Phreatamoeba balamuthi"  
 ORIGIN  
 Query Match 3.5%; Score 85.6; DB 4; Length 951;  
 Best Local Similarity 46.0%; Pred. No. 3.4e-05;  
 Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;  
 Qy 1722 CGCAGATGGCCACCAGCCACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAGNA 1781  
 Db 219 CGCAGATCGCTGGTGTCTGTGCAACATCTTCTCGACGCGAGGACACGCGCGCCGC 278  
 Qy 1782 GATGCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCA 1841  
 Db 279 CATCGCAGCGCGCGTCTGCTTCTGCTGGAAGGGCGGAGAACCTCCAGGAGTACTG 338  
 Qy 1842 GAAGAGACCTGGGAGACCTGGTGACCGACTACTTGGCAGGCGCACCTGGATCCCGAGTG 1901  
 Db 339 GAGGTGACCTGGAAAGGCCCTGTCTTCCGCCCCCTACCAGGCGCTCAGATCATGTCGA 398  
 Qy 1902 GGAGTTCTGAACACCCCGCTGGTGAAGCTGTGTACAGCTGAGAGGAGGCCCAT 1961  
 Db 399 CGACGCGGTGACGCGACTCTGATGATCCACAAGGGGTTCGCGCGGAGGACAAACCCAA 458  
 Qy 1962 CATCGGCGCGAGACCTTCTAGCTGGACGGCGCCCAACCCCGAGACCAAGATCGGCAA 2021  
 Db 459 GCTGCTGAGACGACGAGGGCTCGAGAGGTGCGCTTGCCTCAACAAGCTGCTCAAGCA 518  
 Qy 2022 GCGCGGTACGTGACCGAGCCGCGGCGGAGATCGTGAGCTTGAACGAGACCAACCAA 2081  
 Db 519 GGTCCAGAAGGAGCAGCCCGCTTCTGGCAACAAGATCTCCCGGAGATCCGCGGTGTCA 578  
 Qy 2082 CCAGAGACCGAGCTGACGAGCCATCCAGCTGGCCCTCGAGGACGAGGAGGAGTGA 2141  
 Db 579 CGAGGAGACGACGACTGGCGTNAAGGCTGTACAGCTGTACCGGCGGAGCGGAGTGT 638  
 Qy 2142 CATCGTACCGACAGCCAGTACGCGCTGGGATCATTCAGGCGCCAGCCGACAAAGCGA 2201  
 Db 639 GTTCCCGCGCTCAACGTCAAGAC---TCTNTTCAACAAGCAAGTTGACACATCTA 695  
 Qy 2202 GAGCGAGCTGTGAACCAAGATCATCGACGCTGATCAAGAGGAGAGAGGTGTACTGAG 2261  
 Db 696 GCGGTGCGCCACTCGCTCATCGACGGCATCAAGCGCGCACCGACGTGATCTCGCGCG 755  
 Qy 2262 TGGGTGCGCCCGCAAGAGGCGATCGGCGGCAACGAGCAGATCGACAGCTGGTAGCAA 2321  
 Db 756 CAAGGTGCGCGCTCGTGGCGGCTGAGGCGGCAAGGGCTGCGCGGAGTGTGCGTGG 815  
 Qy 2322 GGGCATCGCAAGGTGTCTTCTCGACGGCATCATTCAGGCGCGGCTGTGATCTACAGTA 2381  
 Db 816 CGGCGAGGCTGCGCGCTCATCTGTGACGGAGATCGACCCCATCTCGGCGCTCAGCGCTC 875  
 Qy 2382 CATGACGACCTGTACTGTTGGGCGAGCGCGCGCTTGGATCGA 2423  
 Db 876 GATGCGCGCTTCGAGGTCACACGCTCGAGGCGGGCTCGA 917  
 RESULT 7  
 CL971508  
 LOCUS  
 DEFINITION  
 OsIFCC021485 Oryza sativa Express Library Oryza sativa (indica  
 CL971508  
 linear GSS 21-SEP-2004







```
Query Match      3.3%; Score 80.2; DB 4; Length 1132;
Best Local Similarity 45.3%; Pred. No. 0.00032;
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

QY 139 CGCGCCCCCGCAGAGGGTGTCTGGAGTGGCGGCAAGGAGGCGCCACCATGAAGAC 198
DB 139 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 79 CGCGCGAGGCAAGACGAGTACCGCGCGGCCANCTGGTGATCCAGGACAAAGCAAG 138
DB 79 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 199 TGCACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGGCC 258
DB 199 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 139 TACACAGCCCAAGTACCGTTCGTGTC-----GTCCGCTTACCAACAGGAGCATGTC 192
DB 139 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 259 CGGAGTTCGCCAGCAGAGCAAGCGGCCCAACAGCCCCACAGCCGCGAGCTGCAGGTG 318
DB 259 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 193 TGCAGATGCGCTACGCCAAGATCGACGGCGACCATCTCTCGCGCGCTACTCGCAC 252
DB 193 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 319 CGCGGCACACCCCGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 378
DB 319 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 253 GAGCTCACCCGCTTCGGCGTCAAGTCGGCTGACCAACTACGCGCGCGCTACGCGACT 312
DB 253 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 379 CAGATCACCTGTGGCAGCGCCCTGTGTAGCATCAAGTGGCGGCGCAGATCAAGGAG 438
DB 379 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 313 GGCCTGTCTGGCGCGCGCGCTGTCTGAGAGCTCAACCTGACTCCAAATACAGAGGT 372
DB 313 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 439 GCCCTGTGTGACACCGCGCGCGCAGACACCGCTGTCTGGAGGAGATGAGCTTCCCGGCAAG 498
DB 439 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 373 GTCAAGAAGGTCAACGGCGGAGGACTACAACTGTGAGGAGCTCGACGAGCGGCCCGGT 432
DB 373 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 499 TGGAGCCCAAGATGATCGCGCGGATCGCGCGGCTTCAATCAAGGTGGCGAGTACAGCAG 558
DB 499 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 433 TTCAAGGCCCTGTCTGAGCTGGCGCTGTGTGCGCACCTCGACTGGCGCGCGGTGTGCCC 492
DB 433 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 559 ATCTGTATCGAGATCTGGCGCAAGAGCCATCGCACCGTGTGATCGGCGCCACCCCG 618
DB 559 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 493 GCCCTCAAGGCGATGTGGGACGGCGGTCAACGTGTCCTCC-----CACAGGAGACCGGTT 549
DB 493 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 619 GTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTTGAATCTTCCCATC 678
DB 619 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 550 GTGCGCTTCAACGGCGACAGAGAGCTCAACCGCGGCTCTCCGCAAGTACATCTTC 609
DB 550 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 679 AGCCCCATCGAGACGCTGCGCGTGAAGTGAAGCGCGCGCATGAGCGGCCCAAGGTGAAG 738
DB 679 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 610 GCGCGCCACGCTGCGCGGTGTATGAAGCTCTCAAGGAGCAGGACGCGCGCGCTTCGAC 669
DB 610 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 739 CAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTCGAGGAGTGGAG 798
DB 739 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 670 CGCCAGTTCTGCGCTACGCCAAGAGGAGGTGTACCGCGGACATGCTCGAGAAGATCTAC 729
DB 670 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 799 AAGGAGGCGAAGATCACCAAGATCGGCGCGCGGAGAACCCCTTACAAACACCCCGTGTGCGC 858
DB 799 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 730 ACCGAGGCCACAGCAGATCCGCGCGCAGCCGACCTTGTCTCCCAAGCGGCTCGNAG 789
DB 730 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 859 ATCAAGAAGAGGAGCAGCACAAGTGGCGCAAGTGTGTGAGATTTCCGCGAGCTGAAACAAG 918
DB 859 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 790 CCCGAGGCGCGCAAGCCCAAGCACTGGGGCAAGCGCAGGCTGACGTACCAAGCGGCAAG 849
DB 790 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 919 CGCACCCAGGACTTCTGGGAGTGCAGTGGGATCCCCACCCCGCC 966
DB 919 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 850 AACCGCGTGGCGCCAGAGAAGGTTCGCTGGGCTACCCCGACGCCCCC 897
DB 850 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
BM320900 1165 bp mRNA linear EST 03-JAN-2002
LOCUS rocketfeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
sequence.
ACCESSION BM320900
VERSION BM320900.1 GI:18055306
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi

Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
1 (bases 1 to 1165)
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
21819461
11830664
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1165 Std Error: 0.00
POLYA-No. Location/Qualifiers
1..1165
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/notes="syn: Phreatamoeba balamuthi"

Query Match      3.2%; Score 79.6; DB 4; Length 1165;
Best Local Similarity 45.1%; Pred. No. 0.00041;
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;

QY 90 CATCAAGTGTCTTCAACTGCGGCAAGGAGGCGCCACATCGCCGCAACTGCCGCGCCCCCG 149
DB 90 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 46 CGTCAAGAACAAAGCGGTACTTCAAGCGCTTCCAGACCCAGTTCGTCCGCGCGGAGGG 105
DB 46 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 150 CAAGAAGCGGTGTGTGAAGTGTGGCAAGAGGAGGCGCACAGATGAAGAGTGCACCGAGCG 209
DB 150 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 106 CAAGACGAGCTACCGCGCGGCCANCTGGTGATCCAGGACAAAGAACAAAGTACAAAGGCC 165
DB 106 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 210 CCAGGCCAACTTCTTCCGCGAGGACCTGGCTTCCCCCAGGCGAAGCGCGGAGTTCCC 269
DB 210 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 166 CAAATACCGTTCGTGTC-----GTCCGCTTACCACAGGGACATCGCTCTCCAGATCGC 219
DB 166 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 270 CAGCGACGAGAACCGCGCCCAACAGCCCCACAGCGCGGAGCTGCAGGTGCGCGCGACAA 329
DB 270 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 220 CTAGCGCAAGATCGAGCGCGACACATCTCTCGCGCGGCTACTCTCGACGAGCTCACCCG 279
DB 220 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 330 CCCCCGAGGAGCGCGCGCGCGAGCGCGAGGCGACCTTGAACCTTCCCGCCAGATCACCCCT 389
DB 330 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 280 CTTGCGGCTCAAGCTCGGCTGACCAACTACCGCGCGGCTTACGCGACTGGCTGCTGCT 339
DB 280 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 390 GTGCGAGCGCCCCCTGTGAGCATCAAGGTGGGCGGCGCAGATCAAGGAGGCGCTGCTGGA 449
DB 390 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 340 GGCGCGCGGTGTGTGAAGAGTCAACCTTACTTCAAGTACGAGGAGTGTCAAGAAGGT 399
DB 340 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 450 CACCGCGCGCGACACACACCGTGTGTGAGGAGATGAGCTGCGCGCAAGTGGAGGCCCAA 509
DB 450 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 400 CAAGCGCGAGGACTTACAACTGTGAGGAGCTCGACGACGGGCGCGCGCTTCAAGGCC-- 457
DB 400 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 510 GATGATCGGCGCATCGCGCGCTTTCATCAAGGTGCGCGAGTACGACCAAGATTCCTGATCGA 569
DB 510 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 458 --TGCTCGAGCGTGGCTGTGCGCACTCGACTGGCGCGCGCTGTTCGCGCGCTCAA 515
DB 458 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 570 GATTCGGCGCAAGAGGCCATCGGCACGTGTGATCGGCGCGCCACCCCGTGAACATCAT 629
DB 570 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 516 GGGCATGTGGAGCGCGCGCGTCAACGTCCCCACAGGAGAGCCCGCTTCGTCGGCTTCAA 575
DB 516 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 630 CGGCGCGCAACATGTGACCCAGCTGGGTGTCACCTTGAACCTTCCCGCATCAGCCCCCATCGA 689
DB 630 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 576 CGGCGCAAGAGGAGGAGTCAACCGCGGCTCTCCGCAAGTACATCTTCGCGGCGCAGGT 635
DB 576 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 690 GACCGTCCCGTGAAGCTGAAGCCCGCATGGACGCGCCCAAGGTGAAGCTAGTGGCCCT 749
DB 690 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Df	636	CGCCGCTACATGAAGCTCTCAAGGAGCAGGACGCCGCCCTTCGACCGCCAGTTCTC	695	QY	330	GTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGGCAGATCAAGAGGCGCCCTGTCTGA	449
QY	750	GACCGAGGAGAGATCAAGGCTCTGACCGCATCTGCGAGGAGATGGAGAAGGAGGCA	809	Df	249	GAAGAAGACCAAACTTCAACCGGGCGCATCTACAGTCCGTGCTGGAGAAGGAGCG	308
Df	696	GGCTACGCGCAAGGAGGTTGTCACCGCCGACATGCTCGAGAAGATCTACACCGGGCCCA	755	QY	450	CACCGGCGCCGACGACACCGTGTGGAGGAGATGAGCCCTGCGCCGCAAGTGGAAAGCCAA	509
QY	810	GATCACCAGATCGGCGCCGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAGAA	869	Df	309	CGCGGGGACTACCTCGGCAAGACGGTGCAGGTGATCCCGCACGTCACCAACGAGATCCA	368
Df	756	-----CAAGCAGATCCGCGCGACCCGACCTTGTCGCCAAGCGGCTCGAAGCCGA	809	QY	510	GATGATCGGC---GGCATCGGGGGTTTCATCAAGGTGGCCAGTACGACAGATCCTGAT	566
QY	870	GGACAGCACCAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAAACAAGCGCACCCAGGA	929	Df	369	GGATATCTCAAGCGCGCGCGGCATCGGCACCGACGCGGTGGGACGTGGCCATCGT	428
Df	810	GGCGCCCAAGCCCAAGCATGGGGCAAGCGCAGGCTGACGTACCAAGCGGCAAGACCG	869	QY	567	CGAGATCTGGCGCAAGAGGCGCATCGGCACCGTGTGATCGGCCGCCACCCCGCTGAACAT	626
QY	930	CTTCTGGAGGTGCACTGGGATCCCGCACCCCGCGCGCTGGAAGAGAGAGAGCGT	989	Df	429	CGAGATCGCGGACCGTGGGGACATCGAATCGCTGCGGTCTCTCGAAGCCGTGGGCA	488
Df	870	CGTCCGCCCAAGAGAGTCCGCTGGGCTACCCCGACGCCGCCCAAGAGCAGTAAATTCC	929	QY	627	CATCGCGCGCAACATCTGTGACCCAGCTGGGTGCAACCTGAACTTCCCATCAGGCCCAT	686
QY	990	GACCGTGTGAGCTGGGCGAGCGCTACTTTCAGCGTGGCCCTGGACGAGG	1039	Df	489	GATGAGCTCGCATGGGCGCGAACTTCGCGCTTCTGTCACCTCACCTACCTGCGCTA	548
Df	930	CGGTGCCCTGTACACGCCCATCTCCGCTGCGGCGTGGCTGCTGCTGCGG	979	QY	687	CGAGACCGTGGCGGTGAAGCTGAAGCCCGGCATGGAACGCCCGCAAGGTGAAGCAGTGGCC	746
RESULT 11				Df	549	CATCGCGCGCGCGGTGAACCTCAAGACCAAGCCCAACCCAGCACACCGGTGCAAGCTGCG	608
CL982770		1680 bp	DNA	QY	747	CTTGACCGGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGG	806
LOCUS				Df	609	CGAGATCGGCATCGACCGCGCGCCCTGCTGTCGCGCGCGGACCGCGCGCATCCCGACA	668
DEFINITION				QY	807	CAAGATCACCAAGATCGGCGCGCGGAAACCCCTTACAAACACCCCGCTGTTGCCATCAAGAA	866
ACCESSION				Df	669	GGAGCGCGCAAGATCTCGCTGTTTCAACCAAGTGGCGGAAATGGGCGGTGATCAGCATGTG	728
VERSION				QY	867	GAAGACAGCACCAGTGGCGGCAAGCTGGTGAATCTTCGCGAGCTGAACAAGCGCACCA	926
KEYWORDS				Df	729	GGAGTGGACACCATCTACAAGGTGCCCGCATGCTGCAACGAGGCGGCTGAGCGGCGCT	788
SOURCE				QY	927	GGACTTCTGGGAGGTGCAGCTGGGCGATCCCGCACCCCGCGCGCTGAAAGAAGAAGAG	986
ORGANISM				Df	789	GATCTGCGACAGCTGCGCTTGAACACCGCGCGCACAGCTTCAAGCGCTGGGACAGGT	848
				QY	987	CGTGACCGTGTGGAGCTGGGCGAGCGCTTCTAGCGTGGCCCTTGGACGAGGACTTCG	1046
				Df	849	CGTGTACGAGACCGAGCATCCGCGGCGAGGTCAACCATCGCATGCTGGGCAAGTACGT	908
				QY	1047	CAAGTACACCGCTTACATCTCCAGCATCAACAACGAGACCCCGCGCATCCGCTACCA	1106
				Df	909	AGACCTGTCCGACAGCTACAACTCGGTGAACGAGCGCTGCGGCCACCGGGCATGAAGAA	968
				QY	1107	GTACAACTGTGTCGCCCGCGGCTGGAAGGCGAGCGCCCGAGCATCTTCCAGAGCAGCATGAC	1166
				Df	969	CCAGTGGCGTGAAGATCGAGCAGCTGATTCGAGACCATCGCGGCGCGGCGGCGG	1028
				QY	1167	CAAGATCTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTGTACTACAGGCCCCCT	1226
				Df	1029	CGAAGACTGGCCAGGTACGACGCGCATCTCTGTGTCGCGCGCTTCCGCGCGCGCGGCT	1088
				QY	1227	GTACGTGGGAGCGACTGGAGATCGGCAGCACCGCGCGCAAGATCGAGGAGCTGGGCA	1286
				Df	1089	GGAGGCAAGATCTGACGCGCGCTACGCGCGCGTACGCGCGCGAGCAGAGGTGCCCTAC	1148
				QY	1287	GCACCTGTGCGCTGGGCTTCCACACCCCGCAAGAAGACACCAAGAGGAGCCCGCTT	1346
				Df	1149	CTGCTGGGATGAGGTGGCCACCATCGAATATCGCGCCACGTGGCGGGGCTTCGCGAA	1208
				QY	1347	CTGCGCATCGAGCTGCAACCCCGCAAGTGGACCGTGCAGGCCATTCGAGCTCCCGGAA	1406
				Df	1209	CGCCCAAGCAGCAATTCGACGCGCGCAGCGCGCATCCCGGTGATCGCGCTGATCACCA	1268
				QY	1407	GGAGAGCTGAGCTGAAACGATCCAGAGCTGTGGGCAAGCTGAAGTGGGCCAGCA	1466
				Df	1269	GTGGAAGGACGCGCGCGCATCCATCCAGACGCGAGAGACCTCCGACCTGGGGGAC	1328
				QY	1467	GATCTACCCCGGCATCAAGGTGCGCGAGCTGTGCAAGCTGTGTCGCGCGCGCAAGGCCCT	1526
Query Match							
Best Local Similarity							
Matches							
612;							
Conservative							
0;							
Mismatches							
844;							
Indels							
6;							
Gaps							
2;							
QY	270	CAGCGAGAGAACCGCGCCCAACAGCCACACGAGCGCGAGCTGCGAGTGGCGGCGCAAA	329	Df	129	CATCAACGTGGACCGCGGACCATGTCCTCCGTTCCAGCACGCGGAGTGTCTCACCA	188
Df	330	CCCCCGCAGCG	389	QY	189	CGACGCGCGCGAGACCGACCTCGGCGCACTACGAGCGTTCATCGAAACCGCAT	248
FEATURES							
source							
1..1680							
/organism="Oryza sativa (indica cultivar-group)"							
/mol_type="genomic DNA"							
/db_xref="taxon:39946"							
/clone_lib="Oryza sativa Express Library"							
/note="Oryza sativa exon trapped genomic sequences"							
ORIGIN							
Query Match							
3.2%;							
Score							
79.6;							
DB							
9;							
Length							
1680;							
Best Local Similarity							
41.9%;							
Pred. No.							
0.00042;							
Matches							
612;							
Conservative							
0;							
Mismatches							
844;							
Indels							
6;							
Gaps							
2;							

```
Db 1329 CATGCGCTGGCGCGCGAAGCTCCGAGCTCGCTCCGCGCACGCTGGCGCACGAGATCTA 1388
Qy 1527 GACCGACATCGTGCCCTCGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAGACCGGAGAT 1586
Db 1389 CGGCAACATCGTCACCGAGCGGCCCGCCACCGGTACGAGGCGAAGTGAATCTACTCGA 1448
Qy 1587 CCGTGGCGAGCCCGTGCACGCGGTGTACTACGACCCCGAGCAAG---GACCTGGTGGCCGA 1643
Db 1449 CCAGCTCCGCAAGCGCGGCTCGTGTATCTCCGCGCTCACGCGAGCGCAACAGCTCACCGA 1508
Qy 1644 GATCAGAGCAGGCGCGACGACCATGACCTACAGATCTACGAGGAGCCCTTCAAGAA 1703
Db 1509 GATCGTGAAGTGGCGCGCGAGGTGCATCCGTGGTTCATCGCGGTGAGTTCACCGCCGA 1568
Qy 1704 CCTGAAGACCGCGAAGTACGCC 1725
Db 1569 GTTCAAGTCCACCCCGTGGAC 1590

RESULT 12
BM321430
LOCUS
DEFINITION
Mastigamoeba balamuthi 0.1153 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
sequence.
ACCESSION
BM321430
VERSION
BM321430.1 GI:18055836
KEYWORDS
EST.
SOURCE
Mastigamoeba balamuthi
ORGANISM
Mastigamoeba balamuthi
REFERENCE
1 (bases 1 to 867)
AUTHORS
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Senses, C.W.,
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
Philippe, H.
TITLE
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE
21819461
PUBMED
11830664
COMMENT
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockefeller.edu
Insert Length: 867 Std Error: 0.00
POLYA=No.
Location/Qualifiers
1..867
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"

FEATURES
source
1..867
Query Match 3.2%; Score 79; DB 4; Length 867;
Best Local Similarity 45.7%; Pred. No. 0.00051;
Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

Qy 58 CAGCGCAGCACTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGGCGCAAGGAG 117
Db 131 CCGCACAAGATGCGCGAGTGCTCGCGGTATCATCTCTGTCGCCCAACAGTTGAAGTAC 190
Qy 118 GGCCACATCGCCCGCAACTGCGCGCCCGCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAG 177
Db 191 GCGCTGACCCCGCTGAGGTCACTCGATCGTATGATGAGCGGCTGATCAAGATCAGCGGC 250
Qy 178 GAGGGCCACCGATGAAGACTGACCGAGCGCCAGCGCCAACTTCTTCGCGGAGGACCTG 237
Db 251 AAGGTCCGCGCCGACACGCGCACTTCCCGCGGGCTTCATGGACGTCGTCTCGATCGACAAG 310
```

```
Qy 238 GCCTTCCCGCCAGGCGAAGGCGCCGCGAGTTCCCGAGGAGCAAAACCGCGCCAAACAGCCCC 297
Db 311 ACCGACGAGCACTTCCGCGCTCTTACGACACCAAGGCGCGCTTCCAGGCGCACCGCATC 370
Qy 298 ACCAGCGCGAGCTGCGAGGT---GGCGGCGACAACCCCGCAGCGAGCGCGCGCCGAG 354
Db 371 AACTCGACCGAGGCGCAAGTTCAAGCTCGGCAAGGTCCGCGCGTGCAGCTCGGCAACAAG 430
Qy 355 CGCCAGGCGCACCTCTGAAGTTCCCGCCAGATCACCTGTGGCAGCGCCCTTGGTGAAGATC 414
Db 431 GGCATCCCGTACCTGTGGTACCCAGACGCGCGCAGATCCGCTACCCCAACCCCGCATC 490
Qy 415 AAGGTGGCGCGCCAGATCAAGGAGGCGCCCTGTGTGACACCGCGCGCGAGCACACCGTGTG 474
Db 491 AAGGTCAACGACACGCTCAAGATCGACTCGCTCGGCGAAGATCATCGACTTCGTGAGG 550
Qy 475 GAGGATGAGCTGCGCGCGCAAGTGGAGCCCAAGATGATCGGCGGATCGGCGGCTTC 534
Db 551 TTCGAGATCGGCAACCTCGTTCATGATCACTGCGCGAGCGCAACCTTGGCGGCTCGCGCTC 610
Qy 535 ATCAAGGTGCGCCAGTACGACCATGATCTGATCGAGATCTCGGCGCAAGAGCCATCGGC 594
Db 611 ATTGTGCGCGCGAGAGACGAGGCGTCTGTGAGATCATCCAGTCAAGAGACCGCGTC 670
Qy 595 ACCGTGCTGATCGCGCCCGCCACCCCGTGAACATCATCGCGCGCAACATCTGACCCAGCTG 654
Db 671 GGCCACCAAGTTCGCGCGCGCTGACCAACGCTCTGTGATCGGCAAGGCGCACCAAGTCC 730
Qy 655 GGTGTCACCTGAACTTCCCGCATCAGCCCCATCAGACCGCGTGGCGCGTGAAGTGAAGCC 714
Db 731 CTCGTCACTGCTGCGCGCGCGCAAGGCGATCAAGAAGTGCATCATCGAGGAGTTCAGGCG 790
Qy 715 GGCATGACGCGCCCGCCAGGTTGAACGAG 741
Db 791 GGCACCGCGCACAGGACCGAGCAGGAG 817

RESULT 13
CL978463
LOCUS
DEFINITION
OsIFCC031917 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION
CL978463
VERSION
CL978463.1 GI:52411427
KEYWORDS
GSS.
SOURCE
Oryza sativa (indica cultivar-group)
ORGANISM
Oryza sativa (indica cultivar-group)
REFERENCE
1 (bases 1 to 1725)
AUTHORS
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G.K.S., Deng, X.W. and Wang, J.
TITLE
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL
Unpublished (2004)
COMMENT
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
Location/Qualifiers
1..1725
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
```

Db	1036	GACCGGAGCCACGAGAGATGCGCCGATGATCGACGAGAGATCGTGAAGTTGGTACCA	109
Qy	1957	CCCATCATCGCGCCCGAGACCTTTTACGTGGACGGCGCCGACCGACGACCA	2012
Db	1096	GGGTACAAGCAGCACGAGTATTCGTACGCGGATGACGACGACGATGATGACGTCAA	1151
RESULT 14			
LOCUS	CL974989	2031 bp	DNA linear GSS 21-SEP-2004
DEFINITION	OsIFCC042895	Oryza sativa Express Library Oryza sativa (indica cultivar-group)	genomic, genomic survey sequence.
ACCESSION	CL974989		
VERSION	CL974989.1	GI:52404497	
KEYWORDS	GSS.		
SOURCE	Oryza sativa (indica cultivar-group)		
ORGANISM	Oryza sativa (indica cultivar-group)		
REFERENCE			
AUTHORS	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. M. and Wang, J.		
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis		
JOURNAL	Unpublished (2004)		
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.		
FEATURES			
source	1. .2031 /organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_lib="Oryza sativa Express Library" /notes="Oryza sativa exon trapped genomic sequences "		
ORIGIN			
Query Match	3.2%;	Score 78.4;	DB 9; Length 2031;
Best Local Similarity	42.1%;	Pred. No. 0.0007;	
Matches	719; Conservative	0; Mismatches 966;	Indels 21; Gaps 4;
Qy	672	CCCATCAGCCCATCGAGACCGTGGCCGCTGAAGCTGAAGCCCGGATGACGCGCCCAA	731
Db	285	CGCCAGAACCGGCGGCGCCACCCGGAGCGGACCATCTACGACCCAAAGCGGCTCAT	344
Qy	732	GGTGAAGCAGTGCCCTGACCGAGGAGAAGATCAAGCCCTGACCGCCCATCTGCGAGGA	791
Db	345	CGGCGGCGAGTTCTCCGACGCGGAGGTGACGCGGACATGAAGCTGTGCGCTTCGCCGT	404
Qy	792	GATGGAGAGGCGGCATGATCACCAGATCGGCCCGCGAGAACCCCTACACACCCCGCT	851
Db	405	CGTCGACCGGAAACGGCAAGCCGACGTGCGCGCTGCGAGGTGAAGACGGCGACGTGCGCGT	464
Qy	852	GTTCGCGCATCAAGAAAGGACAGCAACCAAGTGGCGCAAGCTGGTGACTTTCGCGGAGCT	911
Db	465	GTTTCAGCCCGAGAGGTGAGCCCATGTTGCTCATCGCGATGAGGAGACGCCGAGGC	524
Qy	912	GAAACAAGCGCACCCAGACATTTCTGGAGGTGACGTGGGATCATCCCCACCCCGCGCGCT	971
Db	525	CTACCTCGGCGAAGGTCAACCGCGCGCTGCTGACCGCTCCGCGCTTACTTTCAACGACGC	584
Qy	972	GAGAGAGAGAGCGCTGACCGTGTGGACGTTGGGCGACGCTACTTTCAGCGTGGCCCT	1031
Db	585	GCAGCGCGACGGCCCAAGGACCGCGCGGTTCATCGCCGGGCTCACCGTGCACCGCATCAT	644

QY 1032 GGACGAGGACTTCGGCAAGTACACCGCTTACCATCCCGAGCATCAACAAACGAGACCCC 1091  
Db 645 CAACGAGCCACACCGCCGCCCATCCCTACCGCATCGAACGAAGGCGCGCCGAGAAGAA 704  
QY 1092 CGGCATCCGCTACAGTACAAAG- ----TGCTGCCCCAGGGCTGGAAGGGCAGCCCCAG 1145  
Db 705 CGTCTCTGTTCTGACCTCGCGCGCGCAGCTTCGAGGTGAGCATCTCTGCCATCGACAA 764  
QY 1146 CATCTTCCAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGAT 1205  
Db 765 CGCGCTGTTCCAGGCTCTTGCCACCAAGCGCGACACCCACTCGCGCGGAGGACTTCGA 824  
QY 1206 CGTGATCTACAGGCCCTCTGTAGTGGCAGCGACTGGAGATCGGCCAGACCGCGC 1265  
Db 825 CCAACGCTCATGGACACTTCGTCAAGGTTCATCCGCGGAAGCAGCGCGCGACATCAC 884  
QY 1266 CAAGATCGAG- --GAGCTGGCAAGCAGCTGCTGCGCTGGGCTTCACCAACCCCGACAA 1322  
Db 885 CGCGGAGCGCGCGCTGGGCAAGTTCGCCCGAGTGGAGCGCGCCAGCGCGGCT 944  
QY 1323 GAAGCACCAAGAGAGCCCTTCTCTGCCATCGAGCTGCACCCCGCAAGTGGACCGT 1382  
Db 945 CAGCAACCAAGCAGAGTGGCGCTCGAGTTCGAGTCCCTGTTCCGCGGCTCGACTTGT 1004  
QY 1383 GCAGCCATCGAGCTGCCCGAAGAGAGAGCTGGACCGTGAACGACATCGAAGAGCTGGT 1442  
Db 1005 GGAGCGCTCTCCCGGCGCGGTTCCGAGGAGCTCAACAGCGACCTCTTCAAGAAGACGAT 1064  
QY 1443 GGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGGCGTGTGAA 1502  
Db 1065 GTGCGGTTGAGGAAGGCCATGCGGAGCGCCGCTGAGCAAGGGCGACATCGACGAGAT 1124  
QY 1503 GCTGCTCGCGCGCAAGGCCCTGACCGACATCTGTCGCCCTGACCGAGGAGCCGAGCT 1562  
Db 1125 CGTCTCTCGTGGCGGCGCAGCACAGGATCCCCAAGTTCGACAGCTGCTCAAGGACTACTT 1184  
QY 1563 GGAGTGGCGCGAAGACCGGAGATCTCGCGCGAGCCCGTGCACGCGGTGTACTACACCC 1622  
Db 1185 CGGTGCAAGAGGCCCAACCGCGCGCTCAACCCCGACGAGCGCTGGCGTACGCGCGCG 1244  
QY 1623 CAGCAAGGACCTGTGTGGCGAGATCCAGAAGCAGGCGCCAGCAGTGGACCTACAGAT 1682  
Db 1245 CGTGAGGCGCAGCATCATCAGCGGCCACGTGACGAGACCGCGCGGCTGATGGCAGCTGAT 1304  
QY 1683 CTACCAAGAGCCCTTCAAGAACTGAAGACCGGCAAGTAGCCGCAAGATGCGCACCCGCA 1742  
Db 1305 CGAGCTCGCGCGCTCACCTCGGCTGGAGACCGCGCGGCTGATGGCAGCTGAT 1364  
QY 1743 CACCAACGACGTGAAGAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGGAGAGCATCGT 1802  
Db 1365 CCGCGCAACACCGTGTGTCGACGAAAGAGACGCGAGGTGTTCAACACGTACAAGGACAA 1424  
QY 1803 GATCTGGGCGAAGACCCCAAGTTCCGCTCGCCATCCAGAGGAGACCTGGAGACCTG 1862  
Db 1425 GCAGACAAACGTTGACCATCAGGTGTTCCGAGGCGAGCGAGCATCACGAGGGAACACG 1484  
QY 1863 GTGACCGGACTACTGCGAGGCCACTTGGATCCCGGAGTGGGAGTCTGTGAACACCCCGCC 1922  
Db 1485 GCTGCTCGGAGTTGCACTTCGCGGCTCGCGGCTCGCGCGCGCGGAGGGG-----CGGC 1538  
QY 1923 CTGTGTGAAGCTGTGTGTAACGCTGAGAAAGAGGCCCATCATCGCGCGCGAGACTTCTTA 1982  
Db 1539 GCAGATCGAGGTGACGTTTCAGGTGACGCGAAGCGCATCTCAGCGTGTGTGC----- 1592  
QY 1983 CGTGAAGCGGCGCCCAACCGGAGACCAAGATCGGCAAGGCGCGGCTACGTGACCGACCG 2042  
Db 1593 CGCCGACAAAGGCCACCGGAGGTTCGAGAAAGATCACCATCTCCGCGGACGACCGCAAGAT 1652  
QY 2043 GGGCGCGGAGAAAGATCGTAGCTGACCGAGACCAACCAAGAACGACGAGCTGCAGGC 2102  
Db 1653 CAGCCAGGAGGAGATCGACCGATGTGTGCGAGGCGGAGGATTCGCCGAGGAGACCG 1712  
QY 2103 CATCCAGCTGGCCCTGCAGACAGCGGCGAGCGAGGTGAACATCGTGACCGGACAGCCAGTA 2162

Db 1713 CCGGCACACGAGGAGCAGGTGAGCGCCCGGAACAGCCTGGAGGCGTACGTCTTACAACATCAA 1772  
QY 2163 CGCCCTGGGCTCATCTCAGGCCAGCCCGACAGAGCGAGAGCGAGCTGTGTGAACACAGAT 2222  
Db 1773 GAACACGCTCGGGCGGCAAGATGGCGGACGCCATGGAGGGCGAGAGAGCAAGGTGGA 1832  
QY 2223 CATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCCCGCCACAGGG 2282  
Db 1833 GGAGGCGGTGAGGGAGCGGTACGAGTGGCTGGACGCAACCCGCGCGCGCAAGGAGGA 1892  
QY 2283 CATCGCGGCAACGACAGATCGACAAAGCTGTGTGAGCAAGGCGATCGGCAAGTGTCTT 2342  
Db 1893 GTACGAGGAGAAAGCTGAGGAGCTGGAGGACGTGTGCAACCCCGTCTATGTCGCGGTCTA 1952  
QY 2343 CTTGGACGCGCATCGATGGCGGCATCG 2368  
Db 1953 CCNAGAGTCCGGCGCGCGCGGCG 1978

RESULT 15  
CL952258  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CL952258 2697 bp DNA linear GSS 21-SEP-2004  
OeJRU000500 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
CL952258  
CL952258.1 GI:52364267  
GSS.  
Oryza sativa (indica cultivar-group)  
Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 2697)  
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
Unpublished (2004)  
Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Classes: exon-trapped.  
Location/Qualifiers  
1..2697  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Express Library"  
/note="Oryza sativa exon trapped genomic sequences"

FEATURES  
source

ORIGIN

Query Match 3.2%; Score 78.2; DB 9; Length 2697;  
Best Local Similarity 43.4%; Pred. NO. 0.00077;  
Matches 680; Conservative 0; Mismatches 863; Indels 24; Gaps 6;

QY 193 AAGGACTCACCGAGCGCGCAGGCAACTTCTTCGCGAGGAGCTGGCTTCCCCCAGGGC 252  
Db 1141 AAGGACTACCGGTGAGGCTGAGCAACGACCGCGAGAGCATCTGTGGGTGGCGGTG 1200  
QY 253 AAGGCGCGGAGTTCCCCAGCGAGCAGAAACCGCGCAACAGCCCCACAGCCCGAGCTG 312  
Db 1201 CACCCCTCGTGTATGTCTCGCGCGAGTTCTAGACGCCATCTCTCAACGCGCTCGAGGTG 1260  
QY 313 CAGGTGCGCGGCGCAACACCCCGCAGCGAGGCGCGCGCGCGAGCGCCAGGCGACCTGAAC 372  
Db 1261 TTCAAGGTGAACAACACCGCGCGGCGCTGGCGGACCCCGTCCCGTACAAAGCTG 1320





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 07:40:05 ; Search time 7856 Seconds  
(without alignments)

11933.846 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463

Sequence: 1 gtcagccaccatggccga.....gggctagcaccgtgaattc 2463

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_ges1:\*

9: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	136.4	5.5	330	9	CL293849
2	98.4	4.0	2886	9	CL967755
3	86.8	3.5	2598	3	AV103647
4	86.6	3.5	1398	9	CL961989
5	85.6	3.5	951	4	BM321451
6	85.4	3.5	2031	9	CL974989
7	84	3.4	1725	9	CL978463
8	80.8	3.3	3069	9	CL973991
9	80.6	3.3	869	7	CK159167
10	80.6	3.3	1941	9	CL971508
11	80.2	3.3	1132	4	BM320864
12	79.6	3.2	1165	4	BM320900
13	79.6	3.2	11691	9	CL962901
14	79	3.2	867	4	BM321430
15	78.6	3.2	2682	9	CL965033
16	78.2	3.2	2853	9	CL974397
17	78	3.2	1485	9	CL970981
18	78	3.2	2313	9	CL982362
19	78	3.2	2454	9	CL975440
20	77.4	3.1	1509	9	CL959255
21	77.4	3.1	2559	9	CL982027
22	76.6	3.1	1550	4	BM321022
23	76.4	3.1	2028	9	CL979437
24	76.2	3.1	1401	9	CL962721

25	76	3.1	1680	9	CL982770	CL982770	OsIFSC049
26	76	3.1	2793	9	CL948321	CL948321	OsIFSB004
27	75	3.0	862	4	BM321023	BM321023	rockefell
28	75	3.0	1734	9	CL948553	CL948553	OsIFSB005
29	74.6	3.0	1410	9	CL947314	CL947314	OsIFSB004
30	74.6	3.0	1989	9	CL980973	CL980973	OsIFCC035
31	74.4	3.0	914	8	BZ568300	pac82-164	BZ568300
32	74	3.0	1290	9	CL972679	OsIFCC023	CL972679
33	74	3.0	2072	3	CR603312	full-leng	CR603312
34	73.2	3.0	853	4	BM321393	rockefell	BM321393
35	73	3.0	2151	9	CL972100	OsIFCC041	CL972100
36	72.8	3.0	586	4	BM587428	170006873	BM587428
37	72.6	2.9	788	6	CB643171	OSUNEBO3L	CB643171
38	72.4	2.9	753	9	CC675888	OGWCO51TH	CC675888
39	72.2	2.9	602	7	CV057146	BNL2463	CV057146
40	72.2	2.9	640	2	BE601575	HVSMH009	BE601575
41	72.2	2.9	688	6	CB648640	OSJNEB12C	CB648640
42	72.2	2.9	764	6	CB651670	OSJNEB16L	CB651670
43	72.2	2.9	766	6	CB642928	OSJNEB03F	CB642928
44	72.2	2.9	809	6	CB641397	OSJNEB01A	CB641397
45	72.2	2.9	841	6	CB651502	OSJNEB16H	CB651502

## ALIGNMENTS

RESULT 1  
LOCUS CL293849/c 330 bp DNA linear GSS 12-FEB-2004  
DEFINITION 02S0349-08A1-C03 UniformMu MUTAIL Library zea mays genomic clone  
ACCESSION 02S0349-08A1-C03, genomic survey sequence.  
VERSION CL293849  
KEYWORDS CL293849.1 GI:42541978  
SOURCE GSS.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 330)  
AUTHORS Latchaw S., Tan, B.-C., Settles, A.M. and McCarty, D.R.  
TITLE Sequence tagged transposon insertions from the UniformMu maize population  
JOURNAL Unpublished (2003)  
COMMENT Contact: Donald R. McCarty  
Plant Molecular and Cellular Biology Program  
University of Florida  
PO 110690 Gainesville, FL 32611-0690, USA  
Tel: 352-392-1928 x322  
Email: drmc@ufl.edu  
Sequence flanking probable Mu insertion site in UniformMu line:  
02S0349-08, Primer set: A  
Class: transposon insertion site.

## FEATURES

source  
1. .330  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="W22 (ACR, bz1-m9)"  
/cultivar="UniformMu"  
/db\_xref="taxon:4577"  
/clone="02S0349-08A1-C03"  
/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sepharose 400 spin columns and cloned into the TOPO PCR4 vector."

## ORIGIN

Query Match 5.5%; Score 136.4; DB 9; Length 330;  
Best Local Similarity 66.4%; Pred. No. 2.2e-14;

Matches	211;	Conservative	0;	Mismatches	106;	Indels	1;	Gaps	1;
Qy	421	GGCGCCGAGATCAAGAGGCGCCCTGCTGGACACCGCGCCGACACACCGCTGCGAGGAG	480						
Db	329	GGGGGCGAGCTGGAAGAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAGAA	270						
Qy	481	ATGAGCTCGCCCGGCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGGCTTCATCAAG	540						
Db	269	ATGACTTTGACAGGAAGATGGAAACCAAAATGATAGGGGAAATTGAGGTTTTATCAA	210						
Qy	541	GTGCGCCAGTACGACAGATCCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTG	600						
Db	209	GTAAGACAGTATGATCAGGTACCGATAGAAATCTGTGGGCATAAGCTATAGGTACGGTA	150						
Qy	601	CTGATCGGCCCCACCCCGTGAAATCATATCGGCGCAACATGCTGACCCAGCTGGGCTGC	660						
Db	149	TTAGTAGACCTACACCTGTCAACATTAATTGGAAGAAATCTGTTGATCTAGATTGG-TGC	91						
Qy	661	ACCTGAACATTCCCATCAGCCCCCATCGAGACCGTGCCCGGTGAAGCTGAAGCCCGGCATG	720						
Db	90	ACCTTAAATTTCCCATTAGTCTCTATTGAAATGCTACCAAGTAAATTTAAGCCAGGAATG	31						
Qy	721	GACGGCCCCAAGGTGAAG	738						
Db	30	GATGGCCCAAAAGTAAAG	13						

RESULT 2	CU967755	2886 bp	DNA	linear	GSS 21-SEP-2004
LOCUS	CU967755				
DEFINITION	OaIFCC015718 <i>Oryza sativa</i> Express Library <i>Oryza sativa</i> (indica cultivar-group) genomic, genomic survey sequence.				
ACCESSION	CU967755				
VERSION	CU967755.1	GI:52390149			
KEYWORDS	GSS.				
SOURCE	<i>Oryza sativa</i> (indica cultivar-group)				
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; <i>Oryza</i> .				
REFERENCE	1. (Bases 1 to 2886) Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G.K.S., Deng, X.W. and Wang, J. An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis				
JOURNAL	Unpublished (2004)				
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.				

FEATURES	source	Location/Qualifiers	Class: exon-trapped.
		1. .2886	
		/organism="Oryza sativa (indica cultivar-group)"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:39946"	
		/clone_lib="Oryza sativa Express Library"	
		/note="Oryza sativa exon trapped genomic sequences "	
ORIGIN			
Query Match		4.0%; Score 98.4; DB 9; Length 2886;	
Best Local Similarity		43.2%; Pred. No. 1.7e-07;	
Matches	709; Conservative	0; Mismatches 906; Indels 27; Gaps 4;	
Qy	189	GATCAAGGACTGCACCGGCCAGGCCCACTTCTTCGCGAGGACCTGGCCTTCCCCCA	248
Dd	219	GGTCATGAGGCCGCCGCCCATCATGCGGCATCGTCTCTGCCAACGGCGGCGCAGGCCGCC	278

[illegible]

Db 1356 CGCGGACCGCGGCTCCGCTCGCTGGCGTGTCTGTACAGAAAGTGGCGGAGAG----- 1410  
Qy 1386 CGTGACGCCATCGAGTCTGCCCGGAGAGAGAGTGGACCGTGAACACATCCAGAGAGCT 1445  
Db 1411 -----AGCAAGAGAGAGAGAGAGAGAGAGTGGACCGTGAACATCCGCTCGCGCT 1460  
Qy 1446 GGTGGGCAAGTGAATCTGGGCGACCGCAGATCTACCCCGGCAATCAAGGTGGCGCCAGCTGTG 1505  
Db 1461 GTTCGACCGCGGAGGACGACGCGGAGACCATCGCGCGCGTGTGACCTCGCGGT 1520  
Qy 1506 CAAGCTGTGCGCGCGCGGAGGCGCTTACCGCATGTGCGCCCTGACCGAGGAGGCGGA 1565  
Db 1521 GAACGTGAAGATGATCAACCGCGGAGGCGCTGCGCATCG-----GSAAGGAGAGCGGCGG 1574  
Qy 1566 GCTGGAGTGGCGGAGAACCGCGAGATCTTGGCGGAGCGCGTGTGACCGCGGTGTACTACGA 1625  
Db 1575 GCGGCTCGGATGGGACCAACATGTACCGCTGACGACGCTGTGGCGGACAGAGAG 1634  
Qy 1626 CCCGAGCAAGACCTGTGGCGGAGATCCAGAGAGGCGGACGACGAGTGGACCTTACCA 1685  
Db 1635 CGAGATAGCGGCGCTCCCAATCGACGAGCTGATCGAGAGGCGGACGCGCTTGGCGCGGT 1694  
Qy 1686 GATCTACGAGGCGCTTACGACCTGAGACCGGCAAGTACCGCAGATGCGCACGCG 1745  
Db 1695 GTTCCCGGAGCACAAAGTACGAGATCGTGAAGCGCTTCCAGACCGGAGCACATCTCGCG 1754  
Qy 1746 CCACACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGCAT 1805  
Db 1755 CATGACCGGAGCGCGGTGACGAGCGCGCGCGCTGAGAGGCGGACATCGGCATCGC 1814  
Qy 1806 CGTGATCTGGGCGCAAGACCGCC 1827  
Db 1815 CGTCGACGACGACGACGAGGCC 1836

## RESULT 3

AY103647  
LOCUS 2598 bp mRNA linear HTC 16-OCT-2002  
DEFINITION Zea mays PC0142084 mRNA sequence.  
ACCESSION AY103647

VERSION AY103647.1 GI:21206725

KEYWORDS HTC.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
Clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 2598)

Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,  
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

Maize Mapping Project/DuPont Consensus Sequences for Design of

Overgo Probes

Unpublished (2002)

2 (bases 1 to 2598)

Coe,E.H.

Direct Submission

Submitted (25-APR-2002) Maize Mapping Project, University of

Missouri, Columbia, MO 65211, USA

If you are interested in getting corresponding physical clones,  
these are publicly available from ZmDB and may be found by BLAST  
searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,  
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the  
maize cDNA sequences is either Virginia Walbot, Stanford or Pat  
Schnable, Iowa State, then clones may be requested from ZmDB:  
www.zmdb.iastate.edu.

Location/Qualifiers

1..2598

/organism="Zea mays"

/mol\_type="mRNA"

/db\_xref="MaizeDB:638378"

/db\_xref="taxon:4577"

/clone\_lib="Maize Mapping Project/DuPont Cornsensus

FEATURES  
source

Library" this sequence is part of a project of EST  
/notes= resulting from the application of public  
contigs to seed Dupont contigs; this resource was  
assembled by Dupont as part of a collaboration for the  
overgo addressing of BACs in conjunction with the Maize  
Mapping Project"

## ORIGIN

Query Match 3.5%; Score 86.8; DB 3; Length 2598;  
Best Local Similarity 42.1%; Pred. No. 2e-05;  
Matches 81; Conservative 0; Mismatches 1102; Indels 15; Gaps 5;  
Qy 296 CCACGACCGCGAGCTGCGGCGGAGCAACCCCGCAGAGCGCGCGCGAGC 355  
Db 51 CCACGACCTCCACCTCTCCACCGCGCGCCCAACCAACACACCGAGCGAGCA 110  
Qy 356 GCAGAGGACCTTGAATTTCCCGCAGATCACCTGTGGAGCGCCCTTGGTAGCATCA 415  
Db 111 TGGCGGCAACGCGGCATCTGGAGAGCGACCGCTGAACTGGGGCGCGCGCGG 170  
Qy 416 AGGTGGGCGCGCAGATCAAGGAGGCGCTGTGGACACCGCGCGAGACACCGTGTGG 475  
Db 171 AGTGGCGGAGGCCACTGGAAGGTGAAGCGCATGTGTGGCAGCGCGCGAGCCG 230  
Qy 476 AGGAGATGAGCTTCCCGGCAAGTGAAGGCCCAAGATGATCGGCGGATCGCGGCTTCA 535  
Db 231 TGGTCAAGATCGAGGGTTCCACCTCGCGTGGCGCAGGTGGCGCGCTCGCCTCGCCA 290  
Qy 536 TCAGGTGCGCGAGTACGACAGATCTGTGATCGAGATCTGGGCAAGAGCCATCGGCA 595  
Db 291 AGGACGCGTCCGCGTCCGCGTCCGAGCTCGACGAGGAGGCGCGCGCGCTCAAGGCA 350  
Qy 596 CCGTGTGATCGGCG 655  
Db 351 GCAGGAGTGGATCTCGACTGCTGCCACGCGCGCGCGCATCTACGCGGTCAACACCG 410  
Qy 656 GCTGCACACCTTCCCGCATCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715  
Db 411 GCTTGGCGCGCACCTCCACCGCGCACCAAGGAGCGCGCGCGCGCTCCAGTCCAGCTGC 470  
Qy 716 GCATGAGCGCGCGCGAGGTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 775  
Db 471 TCAGGCACTCTCAACGCGCGGAATCTTGGCGACCGCGCGCGCGCGCGCGCGCGCG 530  
Qy 776 CCGCATCTCGGAGGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 835  
Db 531 AGGTACCG 590  
Qy 836 CCTACACACCG 895  
Db 591 TCCGCTTCGAGATCTTCGAGGCGCATCAGAAAGTGTCTCAACACCGCGGTTCAGCCCTGCC 650  
Qy 896 TGGACTTCGCGGAGTGAACAAGGCGACCCAGGACTTTCGGGAGGTTCAGGTGGGCGATCC 955  
Db 651 TGGCGGTTCGCGGCG 710  
Qy 956 CCCACCG 1015  
Db 711 GCCTCATCAGCG 767  
Qy 1016 ACTTCAGCGTGGCGCGGAGGACTTTCGCGAAGTACACCGCGCTTCAACCATCCCGAGCA 1075  
Db 768 ACGCCCGGAGGGTTCAAGATCGCGGCATCGAGGCGCGGTCTTCAAGCTCAACCCCA 827  
Qy 1076 TCAACAAAGAGACCGCGCGCATCGCTPACAGTACAAAGTGTGCGCGCGCGCGCGCG 1135  
Db 828 AGGAGGCGCTTCGCGCATCGTCAAGCGCACGTCCGCGCGCGCGCGCGCGCGCGCGCG 887  
Qy 1136 GCAGCG 1195  
Db 888 TGTACGAGCGCAACGCTCTGGCGGTCTGTGCGAGGTCCT---GTCCGCGCTCTTTCGCG 944



QY 481 ATGAGCTGCGCCGCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCAAG 540  
 |||||  
 Db 631 AGGATCTCGTCCGTCACAGCCATGTCGCGCTCTGCGAGGCCACCGCGCGGATCGCC 690  
 |||||  
 QY 541 GTGCGCAGTACGACACCATCTGATCGA---GATCTGCGGACAGAGGCGATCGGCACC 597  
 |||||  
 Db 691 GAGTCCGCTACGCGGTGGGCAAGGATCCAGGATCGCGCAAGTTTCTCAACGCCAGC 750  
 |||||  
 QY 598 GTGCTGATCGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGC 657  
 |||||  
 Db 751 GTGGGATCGGTGCTCTGCTTCAGAGACATCTCAACCTGCTG----- 798  
 |||||  
 QY 658 TGCACCTGAACTTCCCATCAGCCCATCGAGACCCTGCCGCTGAAGCTGAAGCCCGGC 717  
 |||||  
 Db 799 TACATCTGCGAGTGCATGGCTCTCCCGAGTGCCCAACTACTGGAAGCAGGTTATCAAG 858  
 |||||  
 QY 718 ATGACGCGCCCAAGTGAAGCATGTGGCCCTGACCGAGGAGATCAAGGCCCTGACC 777  
 |||||  
 Db 859 ATCAACGACTACCAAGAGAGCCGTTCTGTGAACCGCGTGTGCTCCATGTTCAACACC 918  
 |||||  
 QY 778 GCCATCTGCGAGGATGGAGAGGAGGCAAGATCAACAGATCGGCGCCCGAGAACCCC 837  
 |||||  
 Db 919 GTGCCCGCAAGATGCGCGTCTGCGCTTCAGAGAGACACCGGCGACACC 978  
 |||||  
 QY 838 TACAACACCCCGTGTTCGCCATCAAGAAGAGACACACCAAGTGGCGCAAGCTGGTG 897  
 |||||  
 Db 979 AGGAGAGCGCCGCATCGACGCTCTGCAAGGGGCTGATCGSCGCAAGGCGCAAGTTC 1038  
 |||||  
 QY 898 GACTTCGCGAGCTGAACAGCGCACCGAGACTTCTGGGAGGTGAGCTGGGATCCCC 957  
 |||||  
 Db 1039 ATCTACGACCCCAAGTGTGACGAGGACCAAGTTCAGCGCGGACCTTCGCATGAGCAAGTTC 1098  
 |||||  
 QY 958 CACCCCGCGCGCTGAAGAAGAGAGCTGACGCTGTGAGCTGGCGGCGCGCTAC 1017  
 |||||  
 Db 1099 GATTGGATCACCCGCTCATCTCCAGCGATGAGCCCGACGCGCATCAAGAGGTGTCC 1158  
 |||||  
 QY 1018 TTACGCTGCGCTTGGACGAGGACTTCGCGAAGTACACCGCTTTCACCATCCCCAGCATC 1077  
 |||||  
 Db 1159 GTGCGCTGGAGCGCTAGGAGCGCGCAGGACGCCCGCGGCTGTCATCTCAACGAG 1218  
 |||||  
 QY 1078 AACACGAGACCCCGCGATCCGCTAC 1104  
 |||||  
 Db 1219 TGGACGAGTTTCAGGAGCTCGACTAC 1245  
 |||||

## RESULT 5

BM321451

LOCUS

DEFINITION rockefeller.0.1211 Mastigamoeba balamuthi lambda ZAP II Library  
 Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC 3.3.1.1), mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Muller Miklos  
 Laboratory of Biochemical Parasitology  
 The Rockefeller University  
 1230 York Avenue, New York, NY 10021, USA  
 Email: mmuller@rockefeller.edu  
 Insert Length: 951 Std Error: 0.00

POLYA=No.

FEATURES

source

Location/Qualifiers

1..951

/organism="Mastigamoeba balamuthi"

/mol\_type="mRNA"

/strain="ATCC 30984"

/db\_xref="taxon:108607"

/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"

/note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match

Best Local Similarity 3.58; Score 85.6; DB 4; Length 951;

Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

RESULT 6

CL974989

LOCUS

DEFINITION

OsifCC042895 Oryza sativa Express Library Oryza sativa (indica

cultivar-group) genomic, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

Oryza sativa (indica cultivar-group)

CL974989 2031 bp DNA linear GSS 21-SEP-2004

OsifCC042895 Oryza sativa Express Library Oryza sativa (indica

cultivar-group) genomic, genomic survey sequence.

CL974989

CL974989.1 GI:52404497

GSS.

KEYWORDS

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

ORGANISM *Oryza sativa* (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzoae; Oryza.

REFERENCE  
AUTHORS 1 (bases 1 to 2031)  
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,  
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,  
Wong, G. K. S., Deng, X. W. and Wang, J.  
TITLE An analysis of transcriptional regulation of the rice genome and  
its comparison to Arabidopsis  
JOURNAL Unpublished (2004)  
COMMENT Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-8048676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.

FEATURES  
source  
location/Qualifiers  
1..2031  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Express Library"  
/note="Oryza sativa exon trapped genomic sequences "

ORIGIN  
Query Match 3.5%; Score 85.4; DB 9; Length 2031;  
Best Local Similarity 42.6%; Pred. No. 3.6e-05;  
Matches 787; Conservative 0; Mismatches 1026; Indels 36; Gaps 5;  
576 CGGCAAGAGCGCATCGGCACCGTCTGATCGGCCCCACCCCGTGAACATCATCGGCG 635  
|||||  
153 CGGCACGAGTACTCGTGGTGGAGTCTACCGGAACGGCCATGTCGAGATCATCGCAA 212  
|||||  
636 CAACATGCTGACCGAGCTGGGTGACCCCTGAACCTTCCCATCAGCCCATCGAGACCGT 695  
|||||  
213 CACACAGGGGAACCGGATACGCCCTCTGTGGTGGCTTTCACCGACGGCGGCGGCT 272  
|||||  
696 GCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAGGTGAAGTGAAGTGGCCCTGACCGA 755  
|||||  
273 CATCGTGGCGCGCAAGAACCGAGCGGCGGCCAACCGGAGCGGACCATCTACAGCG 332  
|||||  
756 GGAGAAAGATCAAGGCCCTGACCGCCATCTGCGAGAGATGGAAGAGGAGGCAAGATCAC 815  
|||||  
333 CAAGCGGCTCATCGGCGCGGAGTTCTCGGACGCCGAGGTGACGCGCGACATGAAGTCT 392  
|||||  
816 CAAGATCGGCCCGAGAACCCCTACACACCCCGTGTTCGCCATCAAGAGAGGACAG 875  
|||||  
393 GCCGTTCGCGCTGTCGACCGGAAACGGCAAGCCGACGTCGCGCTCGAGGTGAAGGACG 452  
|||||  
876 CACCAAGTGGCGCAAGCTGGTGGATTCGCGAGCTGAACAGCGCACCCAGGACTTCTG 935  
|||||  
453 CGACGTGG-----CGTGTTCAGCCCGGAGGAGGTGAGCGCCATGGTCTCAC 500  
|||||  
936 GAGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTGAAGAGAGAGAGAGCGTGAACCGT 995  
|||||  
501 GCGGATGAAGGAGAGCGGCGGAGGCTTCTCGGCGAGAGAGGTTCACCCCGCGCGTCTCAC 560  
|||||  
996 GCTGGACGTGGCGAGCGCTACTTCAGCGTGGCCCTGGAGGAGGACTTCCGCAAGTACAC 1055  
|||||  
561 GCTCCCGGCTACTTTCAGAGAGCGGAGCGGCGGAGGCGCACCAAGGACGCCGGGTCTATCG 620  
|||||  
1056 CGCCTTTCACCATCCCGAGCATCAACAAAGAGACCCCGGCATCCGCTACCAAGTACAAGT 1115  
|||||  
621 CGGGCTCACCGTCGACCGCATCATCAACGAGCCCAACCGCGCGCATCGCCTACGGCAT 680  
|||||  
1116 GCTGCCCGAGGCTGGAGGGGAGGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCT 1175  
|||||  
681 CGACAAGAGGGCGCGGAGAGAGAGTCTCTGCTCTTCGACCTCGCGCGGCGGACGCTTCCA 740  
|||||

QY 1176 GGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATCTTACCGAGCCGCCCTGTAGCTG-- 1233  
Db |||||  
741 CGTCAGCATCTCGCATCGACACGGGTGTTTCGAGGTCTTTCGCCACCAAGCGGACAC 800  
|||  
QY 1234 ----GGCAGCGACTCTGGAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCA 1289  
Db |||||  
801 CCACCTCGGCGCGGAGGACTTCGACCAACGCTCATGGACCACTTCGTCAAGGTCTATCG 860  
|||  
QY 1290 CTTGCTGGCTGGGGCTTACACACCCCGACAAAGACACCAAGAGAGAGCCCTTCTCT 1349  
Db |||||  
861 CCGAAGCAGCGCGCGACATCACCGCGCGCGCGCTGCGCAAGCTCCGCGGCA 920  
|||  
QY 1350 GTGGATGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCGCATTCGAGCT---GCC 1406  
Db |||||  
921 GTGCGAGCGCCCAAGCGCGCTTCAGCAACGACGACGAGTGGCTGCGGTGAGTGCATC 980  
|||  
QY 1407 CGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAATGGGC 1466  
Db |||||  
981 CTTGTTGAGCGGCTCGACTTGTTCGAGCGCTCTCCCGCGCGGTTTCGAGGAGTCA 1040  
|||  
QY 1467 CAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTGAAGTGTGCGGGGCCCA 1526  
Db |||||  
1041 CAGCGACCTCTTCAAGAAGACGATGGTCCCGTGAAGAAAGCCATGGCGGACGCCGCT 1100  
|||  
QY 1527 GGCCCTGACGACATCGTGCCCTGACCGAGAGCGCGAGCTGGAGCTGGCGCGAGAACG 1586  
Db |||||  
1101 GAGCAAGGCGACATCGACGAGATCGTCTCTGTCGGCGGACGACCAAGATCCCAAGT 1160  
|||  
QY 1587 CGAGATCTGCGCGAGCCCGTGCACCGCGTGTACTTACGACCCCGCAAGGACCTGGTGGC 1646  
Db |||||  
1161 CGACGAGCTGCTCAAGGACTACTTTCGTTGGCAAGAGAGCCCAACCGCGCGCTCAACCCG 1220  
|||  
QY 1647 CGAGATCAGAAGCAGGGCCACGACCATGACCTACCATCTTACAGGAGCCCTTCAA 1706  
Db |||||  
1221 CGAGCGCTGGCGTACGCGCGCGCTGACGCGCCAGCATCATCAGCGGCCAGCTGCAAGA 1280  
|||  
QY 1707 GAACCTGAAGACCGGCAAGTACGCGCAAGTGCACCGCCACCGCCACCAACGACGTAAGCA 1766  
Db |||||  
1281 GAACAGGAGAGATGATCTCTTCGAGCTCGCGCGCTCACCTTCGCGCTGGAGACGGC 1340  
|||  
QY 1767 GCTGACCGAGCCGTGCAAGAGATCGGCATCGAGAGCATCTGTGATCTGGGGCAAGACCCC 1826  
Db |||||  
1341 CGCGCGCTGATGCGCAAGCTGATCCCGCGCAACACGCTGCTGCGACAGAAAGACGCA 1400  
|||  
QY 1827 CAAGTTCGCGCTGCCATCCAGAAGGAGACTGGGAGACCTGGTGAGCCGCTACTTGCA 1886  
Db |||||  
1401 GGTGTTCAACGCTACAGAGCAAGACAAACGCTGACCATCCAGGTGTTTCGAGGGCGA 1460  
|||  
QY 1887 GGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGTGTGAAGCTGTGTA 1946  
Db |||||  
1461 GCGGAGCATGACGAGGAGCAACCGGTCTCTCGGAGGTTTCGNCCTCGCGCGCATCGCGC 1520  
|||  
QY 1947 CCAGTGGAGAGAGGCCCATCATCGCGCCGAGACTTTCAGTGGAGCGGCGCGCA 2006  
Db |||||  
1521 GGCGCGAGAGGGGCGCGCAGATCGAGTGACG-----TTTCGAGGTGGAGCGCA 1571  
|||  
QY 2007 CCGCGAGACCAAGATCGCAAGGCGGCTACGTGACCGACCGGGCCCGGCAAGATCGT 2066  
Db |||||  
1572 CGGCATCTCAGCGTGTGGCGCGCGCAAGGGCCACCGGCGAGGTTCGGAAGATCACCAT 1631  
|||  
QY 2067 GAGCTGACCGAGACCAACCAAGAACCGAGCTGCAGGCAATCCAGCTGGCCCTGCA 2126  
Db |||||  
1632 CTCGGCGACGACCGCAAGATCAGCCAGAGAGATCGACCGGATGGTTCGCGAGCGGA 1691  
|||  
QY 2127 GGACAGCGGCGAGGAGTGAACATCGTGACCGACAGCCAGTACGCGCTTGGGATCATCA 2186  
Db |||||  
1692 GGAGTTCGCGAGGAGGACCGCGGACAGGAGCAGGTGAGCGCCCGCAACAGCTTGA 1751  
|||  
QY 2187 GGCCCGCCCGACAGAGCGGAGCTGGTGAACCGAT-----CATCGACAGCT 2240  
Db |||||  
1752 GGCGTACGCTACCAACATCAAGAACACGCTCGCGGCAAGATGGCGGACCGCATGGAGG 1811  
|||  
QY 2241 GATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCGCGCCCGCCACCAAGGCAATCGCGGCA 2300  
|||

1812 CGAGGAGAGGACAAAGGTGGAGGAGCGGTGAGGAGCGGTACAGTGGTGGACGCCAA 1871  
 2301 CGAGCAGATCGACAAGCTGGTGGACCAAGGGCATCCGCAAGGTGCTGTCTCTGGACGGCAT 2360  
 1872 CCGGACCGCCGCAAGGAGGAGTACGAGGAGAGCTGAGGAGCTGGAGGACGTGTGCAA 1931  
 2361 CGATGGCGGATCGTATCTACCAAGTACATGAGCAGCTGTACGTGGC 2409  
 1932 CCGCGTATGTGCGCGGTCTACCAAGGTCCGGCGGCGGCGCGCGC 1980

## RESULT 7

CL978463 1725 bp DNA linear GSS 21-SEP-2004  
 OaIFCC031917 Oryza sativa Expressed Library Oryza sativa (indica  
 cultivar-group) genomic, genomic survey sequence.

ACCESSION CL978463

VERSION CL978463.1

KEYWORDS GI:52411427

SOURCE GSS.

ORGANISM Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 1725)

Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,

Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,

Wong, G.K.S., Deng, X.W., and Wang, J.

An analysis of transcriptional regulation of the rice genome and

its comparison to Arabidopsis

Unpublished (2004)

Contact: Chen Chen

Department of Bioinformatic

Beijing Institute of Genomics

Chinese Academy of Sciences, Beijing 101300, China

Tel: 86-10-80481559

Fax: 86-10-80488676

Email: chenchen@genomics.org.cn

Rice genomic sequence.

Class: exon-trapped

Location/Qualifiers

1..1725

/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="genomic DNA"

/db\_xref="taxon:39946"

/clone\_lib="Oryza sativa Expressed Library"

/note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Query Match 3.4%; Score 84; DB 9; Length 1725;

Best Local Similarity 43.7%; Pred. No. 6.3e-05;

Matches 470; Conservative 0; Mismatches 600; Indels 6; Gaps 2;

949 GGCATCCCCCAGCCCGCCCTGAAGAAGAGAGCGTGAACCGTGGAGCGTGGC 1008

76 GTCCTCGCAGCCCGGAGCGCTCCACCGCATGTACCGGAGGTGAGCCTCTCGCGAC 135

1009 GAGCCTACTTACGCTGGCCCTGGAGAGAGCTTCGCGAAGTACACCGCTTACCATC 1068

136 CTCGCGCAGCAGCCATCATCGCTTCCAGCGCTCTGGGTCCACCCCTCCGCGCAC 195

1069 CCAGATCAACAGCAGAGACCCCGGCATCCGCTACCAAGTACAGTGGCCCGGC 1128

196 TTCACTTATACCGAGCTCTTCTCTCGGACCTCTCGCTTACCGCTTCCGCTAC 255

1129 TGAAGGCGAGCCCGAGCATCTTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTCCGC 1188

256 CCGCGCTCAGCGCGCGCGTCCGCGCTGGCGCGCGCCATCTCTCCACCGCTCGCC 315

1189 GCCGGAACCCCGAGATCGTGTATCTACAGGCCCTCTGATCGTGGCGAGCCTGGAG 1248

316 TACCTCACTCCGCGCGCTCATCCACCGGACCTCAAGTGCAGACATCTTGTCAAC 375

1249 ATCGGCCAGACCCGCGCAAGATCGAGAGCTCGCAAGCAGCTGCTGCGTGGGGCTTC 1308  
 376 GGCACCTCGGCCAGGTCAAGATCGGCGAAGCTCGGCTCGCGCGCTCTCGCGGCTGC 435  
 1309 ACCACCCCGACA---AGAAGCACCAGAAGAGGCCCCCTTCTGTGTGGTGGCTCAGAG 1365  
 436 ACCTCGCCCGCAGACGCTCATCGGACCGCCGAGTTATGCCCCCGAGATGTACGAGAG 495  
 1366 CTGCACCCCGACAAAGTGGAGCGGTGAGCCCATCGAGCTGCCCGGAGAGAGAGTGGACC 1425  
 496 TGCTACGGCTCGGCTGAGCTTCTCTTGGCATGTGCTGCTCGAGATGCTCACC 555  
 1426 GTGAACGACATCCAGAAGCTGGTGGCAAGCTGAATGGGCCAGCCAGATCTACCCCGGC 1485  
 556 AACGAGTACCCCTACAGCAGGTGCGACAACCGCGCAGATCTACAAGAAGGTCAACGCC 615  
 1486 ATCAAGTGGCCAGCTGTGCAAGCTGCTGGCGGCGCCCAAGGCCCTGACCGACATCG-- 1543  
 616 GGCAAGCTCCCGACGCTTCTACCTCTCACCGACCGCCCGCCGCTTCTCATCGGC 675  
 1544 -TGCCCTGACCGAGGAGCGGAGCTGGAGCTGCCCGAGAACCGCGAGATCTCTCGCGAG 1602  
 676 CGGTGCTTGTGACGCGGCCACCGCCCTCTGCCGAGAGCTCTGCTCGACCCCTTC 735  
 1603 CCCGTGACGCGCTGTACTACGACCCCGACGAGACCTGGTGGCCGAGATCCAGAAGCAG 1662  
 736 CTCTCCCGCTCAGAACACGAGCATCAATACCATTTGCTCATGCCACTTGGCGCGCA 795  
 1663 GGCACGACGAGTGGACCTTACAGATCTACAGAGGCCCTTCAAGAACCTTGAAGACCGGC 1722  
 796 CCGCCATTGCTTGGCGTGCAGTAATTCAGTGAAGAGCAAGAGGAGGAGGCGCGC 855  
 1723 AAGTACGCCAAGATGCGCACCGCCACACCAAGAGCTGAAGAGCTGACCGAGCGCGTG 1782  
 856 GCAGCAAGAGCAGCGGCTATGGCGATCCCGGAGCTGAACAGAGAGCAGCACCATC 915  
 1783 CAGAAGATCGCCATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCC 1842  
 916 TTCTCAAGTGCAGATCGCGCGCGGCAACGTGGGAAACATCTACTTCTCCGTTGAC 975  
 1843 ATCAGAGAGAGACTGGGAGACCTGGTGAACCGACTTGGCGAGGCCACCTGTATGCC 1902  
 976 GTGGCGAACACACGCGATGGAGTGGCGAGATGGTGAAGAGCTGGACATCGCC 1035  
 1903 GAGTGGAGTTGTTGAACACCCCGCTGTAAGCTGTGTACAGCTGGAGTGGAGAGAGAG 1962  
 1036 GACCGGAGCCCGAGAGATTTGCGCATGATGAGCAGGAGATCTGTAGTTGGTACCA 1095  
 1963 CCATCATCGCGCGAGACCTTCTACGTGAGCGCGCCGCCAACCCGAGACCAA 2018  
 1096 GGGTACAGCAGCAGCATGATTCTGTAGCGGATGACGACGATGATGATGATCA 1151

## RESULT 8

CL973991

LOCUS

DEFINITION

OaIFCC05416 Oryza sativa Expressed Library Oryza sativa (indica

cultivar-group) genomic, genomic survey sequence.

CL973991

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (indica cultivar-group)

Oryza sativa (indica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 3069)

Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,

Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,

Wong, G.K.S., Deng, X.W., and Wang, J.

An analysis of transcriptional regulation of the rice genome and

its comparison to Arabidopsis





AUTHORS	Allard, F., Crosby, W. L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D., Genswein, B., Graf, R., Gulick, P., Hrycan, L. D., Laroche, A., Links, M. G., McCarthy, E. L., Monroy, A., Muzak, I., Nilsson, D., Penniket, C., Roach, J. L. and Sarhan, F.
TITLE	Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL	Unpublished (2003)
COMMENT	Contact: Wm L Crosby Bioinformatics University of Saskatchewan, Department of Computer Science 1C101 Engineering Building, 57 Campus Drive, Saskatoon, Saskatchewan, S7N 5A9, Canada Tel: 306 966 1769 Fax: 306 966 2033 Email: fgas_estecs@usask.ca This sequence is the direct result of the Base calling software Phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [128, 636]. Plate: taT537 row: N column: 23.

```

FEATURES
  source
    1. 869
      Location/Qualifiers
        plate: tac337; row: N; column: 23.
        /organism="Triticum aestivum"
        /mol_type="mRNA"
        /cultiivar="Wheat line PI 178383"
        /db_xref="taxon:4565"
        /lab_host="DH5 alpha"
        /clone_lib="Triticum aestivum FGAS: Talt5"
        /notes="Organ: Crown; Vector: pGENE-T; SSH (suppression
        subtractive hybridization) cDNA library from genotype
        PI178383 cold hardened at 2 C for 21 days and 49 days
        (equal amount of cDNA pooled together before subtraction,
        tester) and subtracted against genotype Norstar cold
        hardened at 2 C for 1 day (24 H) (driver). Modified Smart
        cDNA (Clontech) priming and non-directional cloning"

ORIGIN

Query Match          3.3%; Score 80.6; DB 7; Length 869;
Best Local Similarity 44.9%; Fred. No. 0.00025;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

```

	580	AAGAAGGCCATCGGCACCGTGCTGATGGGCCCAACCCCGTGAACATCATCGCCGCAAC	639
Qy			
	827	ATGCGCGCGCCGSCAACAGGAGTCACCAAACAGGACGAGAAACAACCAACACCAC	768
Dd			
	640	ATGCTGACCAGCTGGGCTGCACCTTGAACTTCCCCATCAGCCCCTCGAGACCGTGCCC	699
Qy			
	767	CACAAACGACGACAAACAGCGCACAAACGACCAACAACCAACCAACAACAACAAC	708
Dd			
	700	GTGAAGCTGAAGCCCGGCATCGACGGCCCCAAGGTTGAAGCAGTGGGCCCTTGACCGGAGAG	759
Qy			
	707	ACCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	648
Dd			
	760	AAGATCAAGGCCCTGACCGGCATCTGGGAGGAGATGGAGNAGGAGGCGCATCAACAAG	819
Qy			
	647	AACAACAACGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	588
Dd			
	820	ATCGGCCCGGAGAACCCCTTAACAACCCCCGTGTTCGCCATCAAGAAGAAGGACGACCC	879
Qy			
	587	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC	528
Dd			
	880	AAGTGGCGCAAGCTGGTGGACTTCGGGAGCTGAACAGCGGACACCGAGGACTTCTGGGAG	939
Qy			
	527	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	468
Dd			
	940	GTCGAGCTGGGCATCCGCCACCCCGCGGCTGAAGAAGAAGAAGGCGTGAACCGTGCTG	999
Qy			
	467	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	408
Dd			
	1000	GAGTGGGCGACGGCTACTTTTCAGCGTGCCCTCTGGAAGAGGACTTCGCGCAAGTACACCGCC	1059
Qy			
	407	GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	348
Dd			

Qy	1060	TTCAACATCCCAAGCATCAACAACGAGAGACCCCGGCGATCCGCTTACCAAGTACACAGCTGCTG	1119
Db	347	GACCAACAGCACAACGACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	288
Qy	1120	CCCAGGCTGGAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAG	1179
Db	287	AACA	228
Qy	1180	CCCTTCGCGGCCCGCAACCCCGAGATCGTGATCTACCAAGCCCCCTCTGACGTGGGCAGC	1239
Db	227	AACA	168
Qy	1240	GACCTGGAGATGGCCAGC	1258
Db	167	AACGGCAACAACAACAAC	149
RESULT 10			
LOCUS	CL971508	1941 bp	DNA linear
DEFINITION	OsIFC021485 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.		GSS 21-SEP-2004
ACCESSION	CL971508		
VERSION	CL971508.1	GI:52397596	
KEYWORDS	GSS.		
SOURCE	Oryza sativa (indica cultivar-group)		
ORGANISM	Oryza sativa (indica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.		
REFERENCE	1 (bases 1 to 1941) Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.		
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis		
JOURNAL	Unpublished (2004)		
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.		
FEATURES	Location/Qualifiers		
source	1..1941 /organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_lib="Oryza sativa Express Library" /note="Oryza sativa exon trapped genomic sequences "		
ORIGIN			
Query Match	3.3%;	Score 80.6;	DB 9;
Best Local Similarity	42.4%;	Pred. No. 0.00026;	Length 1941;
Matches	757;	Conservative	0; Mismatches 1004; Indels 24; Gaps 5;
Qy	267	CCCCACGACGAGAACCCGGCCCAACAGCCCCACACCGCCGAGCTCGAGTGGCGGCGA	326
Db	165	CGCGCCAGAACCAAGTCGCCATGAACCAACCAACCGCTTTCGATCGAAGCGGTT	224
Qy	327	CAACCCCCGACGAGCGCGCGCCGAGCGCCAGGCAACCTTGAACTTCCCCAGATCAC	386
Db	225	GATGTGTAGAGGTTCTTCGACCCCGCTCCGTGCAGAGTGATGAACTATGGCCGTTCAA	284
Qy	387	CCTGTGGCAGCGCCCTCTGGTGTAGCATCAAGGTGGCGGCCAGATCAAGAGGCCCTGCT	446
Db	285	GGTGTGCTTGGCCCCCGGCGACAGCCGATGATCTGCTCCAGTACAAGGGCGAGAGAA	344
Qy	447	GGACACCGGCGCGACACCGTGTGTGAGGAGATGAGCTTGTCCCGGCAAGTGGAGCC	506

```
Db 345 GCAGTTCCGCGGAGGAGATCTCGTCCATGCTGCTGATCAAGATGAGGAGATCGCGGA 404
Qy 507 CAAGATGATCGCGGCATCGCGGCTTCAATCAAGGTGCGCCAGTAGACCAAGATCCTGAT 566
Db 405 GGCCTACTCGGCTCGTCCATCAAGAACGCGGTGTCAACCGTCCCGGCTACTTCAACGA 464
Qy 567 CGAGATCTCGCGCAAGAGGCGCATCGGACACGCTGCTGATCGGCGCCACCCCGTGAACAT 626
Db 465 CTCGCAGAGGAGGCCACCAAGGACGCGGCGTCAATCGCGGGCTCAAGTGATCGGAT 524
Qy 627 CATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTGAACTTCCCATCAGCCCAT 686
Db 525 CATCAACGAGCCACCGCGCGCATCGCGTACGCGCTCGACAAAGAGGCGACGAGCAG 584
Qy 687 CGAGACGTCGCGTGAAGCTGAAGCGCGGCGATGAGCGGCCCAAGGTGAAGCATGTGCC 746
Db 585 CGCGGAGAGAACGTGCTCATCTTTGACCTCGCGCGCGGCAGTTCGATGTGTCTGCTCT 644
Qy 747 CTTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGGAGAGGAGGG 806
Db 645 CACCATCAGAGGAGGCACTTCGAGGTCAAGGCCACCGCGCGGCAACCCACTCGCGGG 704
Qy 807 CAAGATCACCAAGATCGGCGCGGAGAACCCCTTAC---AACACCCCGCTGTTGCGCATCAA 863
Db 705 CGAGGACTTCGACAAACCGCATGTGTGAACCACTTGTGCGAGGATTCAAAGCGCAAGACAA 764
Qy 864 GAAGAAGACAGCACCAAGTGGCGCAAGCTGTGTGACTTTCGCGAGCTGAACACGCGAC 923
Db 765 GAAGACATCAGCGGCAACCCGCGCGCTGCGCGGCTCAGGACGGGTGCGAGCGGCG 824
Qy 924 CCAGGACTTCTGGGAGGTGCAGCTGGGATCCCGCACCCCGCGCGCTGMAAGAAAGAA 983
Db 825 CNAAGCGAGCTGTCTGTCGACGCGCGACACCACTTGAAGTCGACTCGTGTACAGGG 884
Qy 984 GAGCGTGACGCTGCTGCGAGCGGCGAGCGCTTACTTACGCGTGCCTTGGACGAGACTT 1043
Db 885 CATCGACTTCTACACGACGATCACAGGCGGGTTCGAGGAGCTCAACATGACCTGTT 944
Qy 1044 CGGGAAGTACACGCGCTTCAACATCCGAGCATCAACACGAGACCCCGGATCCGCTA 1103
Db 945 CGCGAAGTGCATGGAGCGCGCTGGAGAGTGCCCTCGCGACGCGCAAGATGGACAAAGAGCAG 1004
Qy 1104 CCAGTACAAAGCTGCTGCCCGCGGCTGGAAGGGCAGCGCCAGCATCTTCCAGAGCAGCAT 1163
Db 1005 CTGTCAGAGCTGCTGCTGCTCGGCGGCTCACCGCGATCCCAAGGTGCAGCAGCTGCT 1064
Qy 1164 GACCAAGATCTTGGAGCGCTTCCGCGCGCGCAACCCGAGATCGTGTACTACAGGCGCC 1223
Db 1065 CCAGGACTTCTTCAACGGCAAGGAGCTCTGCA-----AGAGCATCAACCCCGACGAGGC 1118
Qy 1224 CTTGTAGTGGGACGAGCTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTGG 1283
Db 1119 CTTGCGGTACGCGCGCGCGCTCCAGCGCGCATCTCTGAGCGCGGAGGCAACGAGAGGT 1178
Qy 1284 CAAGCACCTGCTGCGCTGCGGCTTCAACCCCGGCTTCAACAGAGACCAAGAGGACCC 1343
Db 1179 CGAGACCTGCTTCTGCTGAGCTCACCCCGCTGCTCTTCCGCTTCGAGACCGCGGCGG 1238
Qy 1344 CTTCTGTGGATGGGCTAGAGCTGCACCCCGCAAGTGGACCGTGCA---GCCATCGA 1400
Db 1239 CGTCATGACCGTGTCTATCCCGAGGAACACCACTCCCGACCAAGAGGAGCAGTCTT 1298
Qy 1401 GCTGCCGAGAGGAGGTGACCGTGAACGACATCCGAGAGCTGGTGGGCAAGCTGAA 1460
Db 1299 CTCACCTACTCTCCGACAAACAGCCCGGCTCTGATTCAGAGTGTACGAGGCGAGCGGC 1358
Qy 1461 CTTGGGCCACGAGATCTACCCCGGATCAAGGTGCGCAGCTGTGCA-----AGCT 1511
Db 1359 CAGGACCAAGGACAACTCTCTCGGCAAGTTCAGCTTCTCCGCGATCCCAACCGCGGC 1418
Qy 1512 GTCGCGCGCGCCAGAGGCGCTGACCGACATCGTGCCTTGCACCGAGGAGCGGAGCTGGA 1571
Db 1419 GCGCGGCGTGCCTCAGATCAAGCTCTGCTTTCGACATCGACGCCAACCGGCATCTGAACGT 1478
```

```
Qy 1572 GCTGCCGAGAACCGCGAGATCTTGGCGAGCCGCTGTGACGCGGTACTACGACCCCGAG 1631
Db 1479 GTCCCGGAGGACAAGACGACGGGCGAAGAACAAAGATCAACATCACCACGACGAGGG 1538
Qy 1632 CAAGACCTGTGGCCGAGATCCAGAACGAGGCGCACGACCATGAGGAGCTTACAGATCTA 1691
Db 1539 CGGCTGAGCAAGGAGAGATCGAGAAGATGGTCAGGAGGAGAGTACAAGGCCGA 1598
Qy 1692 CAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCCAAGATGCGGACCGCCACAC 1751
Db 1599 GACGAGGAGCACAAAGAAAGTGGACGCCAAAGACGCCCTCGGAGAACTACGCGCTACAA 1658
Qy 1752 CAACGACGTGAGCAGCTGACCGAGGCGCTGCAGAAGATCGCATGGAGAGCATGCTGAT 1811
Db 1659 CATGCGCAACACCATCAAGGACGAGAAGATCGCTTCAAGTTCGCGCGGACGACAAGAA 1718
Qy 1812 CTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGGAGACCTGGGAGACCTGTGTG 1871
Db 1719 GAGGATCAGAGACGCCATCGATGGGCCATCAGCTGGCTCGACACCAACAGCTCGCGGA 1778
Qy 1872 GACGACTACTGGCAGGCGCACCTGATCCCGAGTGGAGTTCGTGAACACCCCGCCCT 1931
Db 1779 GCGCGACGAGTTCGAGGACA---AGATGAAGAGCTGGAGGGCATCTGCAACCCGATCAT 1835
Qy 1932 GGTGAAGCTGTGTACCAAGCTGGAGAGGACCCATCATCGCGCGGAGACCTTCTACGT 1991
Db 1836 CGCCAAAGATGTACAGGGCCCCGCGCGCATGTGCGGAGGATGGAGAGGAGCGGCC 1895
Qy 1992 GGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCCGCTA 2036
Db 1896 GCGCGCGGACGCGCGGCGAGGCCCAAGATCGAGGAGGTGCGACTA 1940
```

## RESULT 11

```
BM320864
LOCUS
DEFINITION
Mastigamoeba balamuthi 1132 bp mRNA linear EST 03-JAN-2002
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
sequence.
```

## ACCESSION

```
BM320864
BM320864.1 GI:18055270
```

## KEYWORDS

```
EST.
```

## SOURCE

```
Mastigamoeba balamuthi
```

## ORGANISM

```
Mastigamoebidae; Mastigamoeba.
```

## REFERENCE

```
1 (bases 1 to 1132)
```

## AUTHORS

```
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
```

```
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
```

```
Philippe,H.
```

```
The analysis of 100 genes supports the grouping of three highly
```

```
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
```

```
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
```

```
21819461
```

```
11830664
```

```
Contact: Muller Miklos
```

```
Laboratory of Biochemical Parasitology
```

```
The Rockefeller University
```

```
1230 York Avenue, New York, NY 10021, USA
```

```
Email: mmuller@rockvax.rockefeller.edu
```

```
Insert Length: 1132 Std Error: 0.00
```

```
POLYA=No.
```

```
Location/Qualifiers
```

```
1..1132
```

```
/organism="Mastigamoeba balamuthi"
```

```
/mol_type="mRNA"
```

```
/strain="ATCC 30984"
```

```
/db_xref="taxon:108607"
```

```
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
```

```
/note="syn: Phreatamoeba balamuthi"
```

Query Match

3.3%; Score 80.2; DB 4; Length 1132;

Best Local Similarity 45.3%; Pred. No. 0.0003;  
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

Qy 139 CGCGCCCCCGAAGAGGGTGTGGAAAGTGGCGAAGGAGGGCCACAGATGAAGGAC 198  
Db 79 CGCGCGAGGCGAAGAGCGACTACCGCGCGCCANCTGGTATCCAGGACAGACAG 138  
Qy 199 TGCACCGAGCGCAGGCCAACTTCTTCGCGAGAGACTGGCGCTTCCCGCCAGAGGCC 258  
Db 139 TACAACAGCCCAAGTACCGCTTGGTC-----GTCCGCTTCAACAAAGGAGCATCGTC 192  
Qy 259 CGCGAGTTCGCCAGGAGCAGACCGCGCCACAGCCCCACAGCGCGGAGCTGCAGGTG 318  
Db 193 TGCCAGATCGCTACGCCAAGATCGAGCGGACCAATCTCTCGCGCGCGCTACTCGCAC 252  
Qy 319 CGCGCGACAAACCCCGCAGCGAGCGCGCGCGCGAGCGCGGACCGCTGAATTTCCCC 378  
Db 253 GAGCTCACCCCGTGGGGGTCAAGCTCGGCTTGACCACTACGCGCGCGCTACGCGACT 312  
Qy 379 CAGATCACCTGTGGCAGCGCCCGCTGTGTGAGCATCAAGGTGGCGGCGCAGATCAAGGAG 438  
Db 313 GGCCTGTCTGTGGCGCGCGGTGTGTGAAGAAGCTCAACCTCGACTCCAACTACGAGGGT 372  
Qy 439 GCGCTGTGGACACCGCGCGCGCGCGACACCGTGTGTGGAGAGATGAGCTCCCGGCAAG 498  
Db 373 GTCAAGAAGGTCAACCGCGAGGACTTACAACTCGAGGAGCTCGACGACGGCGCGCGGT 432  
Qy 499 TGAAGCGCCAGATGATCGCGGGGATCGCGGGCTTTCATCAAGGTGGCGCAGTACGACCG 558  
Db 433 TTCAAGGCGCTGTCTGAGCTCGGCTGTGTGCGACCTGATGCGGCGCGGGTGTTCGCC 492  
Qy 559 ATCTGTATCGAGATCTCGCGCAAGAAGGCCATCGGCACCGTGTGTATGGCGCCCGCCGCC 618  
Db 493 GCGCTCAAGGCGATGTGGCGCGGGGTCAACGTTCGCC---CACAGCGAGACCGGCTTC 549  
Qy 619 GTGAACATATCGCGCGCAACATGTGACCGAGTGGCTGCGCTGACCTGAACTTCCCGATC 678  
Db 550 GTCGGGTTCACCGCGCAAGAAAGGAGTCAACCGCGGGTCTTCGCGAAGTACATCTTC 609  
Qy 679 AGCCCCATCGAGACCGTCCGCTGAAGCTGAAGCGCGCGCATGGAGCGCGCCCAAGGTGAAG 738  
Db 610 GCGCGCCACGTGCGCGGTATCATGAGCTCTCAAGAGCGAGGAGCGCGCGCTTCGAC 669  
Qy 739 CAGTGGCCCTTGACCGAGGAGAAATCAAGGCCCTTGACCGCATCTGCGAGGAGATGGAG 798  
Db 670 CGCCAGTCTCGCGCTACGCCAAGAGGGTGTACCGCGGACATGCTCGAAGATCTAC 729  
Qy 799 AAGGAGGCAAGATCAACAGATCGGCGCGCGAGAACCGCTTCAACACCGCGGTGTTCGCC 858  
Db 730 ACCGAGGCGCAACAGCAGATCGGCGCGGACCGGACCTTCTGCTCCCAAGCGCGCTCGAAG 789  
Qy 859 ATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGAGTCTTCGCGAGCTGAACAAG 918  
Db 790 CCCGAGGCGCGCAAGCCCAAGCATCTGGGGCAAGCGCAGGCTGAGTACCGAGCGGCAAG 849  
Qy 919 CGCACCCAGGAGTCTTGGAGGTTGAGCTGGGCGATCCCGCCACCGCGCC 966  
Db 850 AACCCTGTGCGCCAGAGAAGGTCCGCTGGGCTACCCCGAGCGCCCC 897

RESULT 12  
BM320900  
LOCUS  
DEFINITION BM320900 1165 bp mRNA linear EST 03-JAN-2002  
rockefeller 0.353 Mastigamoeba balamuthi lambda ZAP II Library  
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA  
sequence.  
ACCESSION BM320900  
VERSION BM320900.1 GI:18055306  
KEYWORDS Mastigamoeba balamuthi  
SOURCE EST.  
ORGANISM Mastigamoeba balamuthi  
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.  
REFERENCE 1 (bases 1 to 1165)

AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,  
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and  
Philippe, H.  
The analysis of 100 genes supports the grouping of three highly  
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba  
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819461  
11830664  
Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockefeller.edu  
Insert Length: 1165 Std Error: 0.00  
POLYA=No.

FEATURES  
source

Location/Qualifiers  
1..1165  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.2%; Score 79.6; DB 4; Length 1165;  
Best Local Similarity 45.1%; Pred. No. 0.00038;  
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;

Qy 90 CATCAAGTGTTCNACTGCGCGAAGAGGGGCGACATCGCGCGCAACTGCGCGCGCCCCCG 149  
Db 46 CGTCAAGAACAAGGCGTACTTCAAGCGCTTTCAGACCCAGTTCCGTCGCGCGCGAGGG 105  
Qy 150 CAAGAAGGGTGTGGAAAGTGGCGCAAGAGGGGCCACAGATGAAGACTGCACCGAGCG 209  
Db 106 CAAGACGAGTACCGCGCGCCANCTGGTGATCCAGGACAAAGATACACAGCCC 165  
Qy 210 CCAGGCGCAACTTCTTCGCGGAGGACCTTCCCGCAGGCGAAGGCCCGCGAGTTCCC 269  
Db 166 CAAGTACCGTTCGTC-----GTCCGCTTCAACCAAGGAGCATCGTCTCCAGATGC 219  
Qy 270 CAGGAGCAGAACCGCGCCCAAGAGCCCAAGCGCGGAGTGCAGGTGCGCGGCGGACAA 329  
Db 220 CTAGCGCAAGATCGACGCGGACCACTCTCGCGCGCGCTTCTCGACGAGCTCACCG 279  
Qy 330 CCGCGCAGGAGCGCGCGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCG 389  
Db 280 CTTGCGGCGTCAAGTCTGGCTGTGACCACTACCGCGCGCGCTTACCGGACTTGGCGCT 339  
Qy 390 GTGCGACGCGCGCTGTGTGAGCATCAAGGTGGGCGCGCGGCGGCGGCGGCGGCGGCGG 449  
Db 340 GCGCGCGGTGTGTGAAGAAGCTCAACCTTCTTCAAGTACGAGGCTGTCAAGAAGGT 399  
Qy 450 CACGCGCGCGCAGACACCGTGTGTGAGGAGATGAGCTTGCCTCGCGGCAAGTGAAGCGCAA 509  
Db 400 CAACGCGGAGGAGTACAAACGTCGAGGAGTCTGACGAGGCGCGCGCGCGCTTCAAGGCG 457  
Qy 510 GATGATCGGCGCATCGCGGCTTTCATCAAGGTGCGCGGAGTACGACGAGTCTCTGATCGA 569  
Db 458 --TGCTCGAGCTGCGCTGTGTGCGCACCTGACGTGGCGCGCGCGCTGTTCGCGCGCTCAA 515  
Qy 570 GATCTGCGCGCAAGAGGCGCATCGCGACCGTGTGTGATCGGCGCGCGCGCGCGCGCGCG 629  
Db 516 GGGCATGTGCGAGCGCGCGCTCAACGTCCTCCCAAGCGAGGCGCGCTTCTGCGGCTCAA 575  
Qy 630 CGGCGCGCAACATGTGTACCGAGTGGGCTGCACTTCCGAACTTCCCGATCAGCCCGCATCGA 689  
Db 576 CGGCGCAACAAGAGGAGTCAACGCGCGGTCTCTCGCAAGTACATCTTTCGCGCGCGCACGT 635  
Qy 690 GACGTCGCGCTGAGCTGAAGCGCGCGCATGGAGCGCGCGCGCGCGCGCGCGCGCGCGCT 749  
Db 636 CGCGCGGTATCAAGAGCTCTCAAGGAGGAGGAGCGCGCGCGCGCGCTTCGACCGCGAGTTCTC 695

Qy	750	GACCGAGGAGAAGATCAAGGCCCTTGA	CGCCATCTCGAGAGATGAGAAAGGAGCA	809		
Db	696	CGCTACGCCCAAGGAGGGTGTCA	CGCGCGCATCTCGAGAAGATCTACACCGAGGCCCA	755		
Qy	810	GATCACCAAGATCGGCCCGGAGACCCCTTAC	MACACCCCGTGTTCGCCATCAAGAGAA	869		
Db	756	-----CAAGCAGATCCGCGCGGACCGACCTT	CGTCCCCAAGCGCGCTTCGAAGACCCCGA	809		
Qy	870	GGACAGCACCAAGTGGCGGCAAGCTGGTGG	AGACTTCCCGCGAGCTGAACAAGCGCACCCAGGA	929		
Db	810	GGGGCCCAAGCCCAAGCACTGGGGCAAGCGG	AGCTGACGTACACAGGCGCAAGACCG	869		
Qy	930	CTTCTGGGAGGTGCAGCTGGGCATFCC	CCCCACCCCGCGCTGAAGAAGAGAACGCGT	989		
Db	870	CGTCGCCCAAGAAGGTTCGCTTGGCTAC	CCCCGACGCCCAAGAGCCAGTAAATTC	929		
Qy	990	GACCGTCTGGAGCTGGGGCAAGCGCTACTT	CAGCGTGCCTCCCTGACACGAGG	1039		
Db	930	CGGTGCCTGTACACCCGCATCTCCCGCT	GGCGGTTGCGCTGCTGCGG	979		
RESULT 13						
CL962901		11691 bp	DNA	linear	GSS 21-SEP-2004	
LOCUS	OsIFCC008219	Oryza sativa	Express library	Oryza sativa	(indica cultivar-group)	genomic, genomic survey sequence.
DEFINITION	CL962901					
ACCESSION	CL962901.1	GI:52380530				
KEYWORDS	GAS.					
SOURCE	Oryza sativa (indica cultivar-group)					
ORGANISM	Oryza sativa (indica cultivar-group)					
	Eukaryota; Viridiplantae; Streptophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.					
REFERENCE	1 (Bases 1 to 11691)					
AUTHORS	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Zhao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G.K.S., Deng, X.W. and Wang, J.					
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis					
JOURNAL	Unpublished (2004)					
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.					
FEATURES	Location/Qualifiers					
source	1..11691 /organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_lib="Oryza sativa Express Library" /note="Oryza sativa exon trapped genomic sequences"					
ORIGIN						
Query Match	3.2%;	Score 79.6;	DB 9;	Length 11691;		
Best Local Similarity	43.5%;	Pred. No. 0.00044;				
Matches 587;	Conservative 0;	Mismatches 739;	Indels 24;	Gaps 4;		
Qy	150	CAAGAGGGCTGTGGAAGTCCGCGCAGGAGGGCCAC	CAGATGAAGACTGACCCGAGCG	209		
Db	2991	CAAGAAAGCCCTTCGACCAAGCAGGCGCTGAA	GGTTCAGTGGCGGTGAAGATCGCGCAAGT	3050		
Qy	210	CCAGGCCAACTTCTTCGCGGAGGACCTGGCCTT	CCCCCAGGGCAAGGCCCGCGAGTTCCC	269		
Db	3051	GAAGACGGGGCGGACGCGGTGACCGTGGCCCTA	CACGATGCCAAGGGCGGCGGCCGGA	3110		
Qy	270	CAGCGAGCAAGACCGCGCCCAACAGCC	CCCCACAGCGCGAGTGCAGTGTGCGCGCGCA	329		

```

Qy 1392 GCCATCGAGCTGCCGGAAGAGGAGCTGACCGTGAACGACATCAGAGCTGGTGGG 1451
Db 4245 GAAGCTGAGGTGGTGAACCTGGACACCGCAGCACTACCGCGCGCGCGCTGGAGCA 4304
Qy 1452 CAAGCTGAACCTGGCGGACGAGACTACCC 1481
Db 4305 GGTCCGGCTTACCACCTGCCGCTGGGCC 4334

RESULT 14
BM321430 867 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
sequence.
BM321430
BM321430.1 GI:18055836
EST.
Mastigamoeba balamuthi
Mastigamoeba balamuthi
Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 867)
AUTHORS Baptiste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockefeller.edu
Insert Length: 867 Std Error: 0.00
POLYA-No.

FEATURES
source Location/Qualifiers
1..867
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match 3.2%; Score 79; DB 4; Length 867;
Best Local Similarity 45.7%; Pred. No. 0.00048;
Matches 314; Conservative 0; Mismatches 370; Indels 3; Gaps 1;

Qy 58 CAGCGCAGCACTTCAAGGGCCCCAAGGGCATCATCAAGTCTTCAACTGGGCAAGGAG 117
Db 131 CCGCACAAGATGCGCGAGTGCTGCCGGTATCATCTCTGTGCGCAACAGGTTGAATAC 190
Qy 118 GGCCACATCGCCCGCAACTGCGGGCCCCCGCAAGAGGCTCTGGAAGTGGGCAAG 177
Db 191 GCGGTGACCGCGCTGAGGTCACTTCGATGTGATGAGCGCTGATCAAGATCGAGCG 250
Qy 178 GAGGGCCACAGATGAAGGACTGCACGCGCGCAGGCCCAACTTTCTTCGCGAGGACCTG 237
Db 251 AAGGTCCGCACCGACACGCACTTCCCCCGGGCTTCATGAGAGTCTGATCGATCGACAAG 310
Qy 238 GCCTTCCCCCAGGGCAGGGCCCGGAGTTCCCCGAGCAGACCGCGCCCAAGCCCC 297
Db 311 ACCGACGAGCACTTCCGCTGCTCTACGACACCAAGGGCGCGTTCAGGGCGCACCGCATC 370
Qy 298 ACCAGCGCGAGCTGCAGGT---GCGCGGCGACAAACCCCGCAGCGAGGCGCGCGCGAG 354
Db 371 AACTCGACCGAGGCAAGTTCAGCTCGGAGAGTTCGCGCGCTGCGAGCTCGGCAACAAG 430
Qy 355 CGCAGGGCAACCTTGAACCTTCCCCCAGATCACCTGTGGCAGCGCCCCCTTGGTGAATC 414

```

```

Db 431 GGCAATCCCGTACCTGGTGAACCCACGACGCGCGACGATCCGCTACCCCAACCCCGACATC 490
Qy 415 AAGTGGCGCGCCAGATCAAGAGGAGCCCTGTGTGACACCGCGCGCGAGCAGCACCGTGTCTG 474
Db 491 AAGGTCAACGACACGGTCAAGATCGACTGGCTCGGGCAAGATCATCGACTTGTCTAGG 550
Qy 475 GAGGAGATGAGCTTGGCCGCAAGTGAAGCCCAAGATGATCGCGCGGATCGGGGGTTC 534
Db 551 TTGAGATCGGCAACCTCGTCAATCATCTGCGGCGGACCACTTGGCGCGTGGGGCTC 610
Qy 535 ATCAAGTGGCGCAGTACGACCAATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGC 594
Db 611 ATTGTGCGCGCGAGAGACGAGGCGTCTGTTGAGATCATCCACGTCGAAGGACGCGGTC 670
Qy 595 ACCTGTCTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTG 654
Db 671 GGCCACCAGTTGCGGACGCGCTGACCAAGCTTCTGATCGGCAAGGGCACCAGTCC 730
Qy 655 GGCTGCACCTGAACATTTCCCATCATGCCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCC 714
Db 731 CTGTCACGCTGCGCGCGCAAGGCGCATCAAGAAGTCGATCATCGAGGAGTTCAGGCG 790
Qy 715 GCATGACACGCGCCCAAGGTGAAGCAG 741
Db 791 CGCCACGCGCACAGGACCGAGCAGGAG 817

RESULT 15
CL969033 2682 bp DNA linear GSS 21-SEP-2004
LOCUS OsIFCC017680 Oryza sativa Express Library Oryza sativa (indica
cultivar-group) genomic, genomic survey sequence.
CL969033
CL969033.1 GI:52392695
GSS.
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 2682)
Ma,L., Wang,C., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
Location/Qualifiers
1..2682
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN
Query Match 3.2%; Score 78.6; DB 9; Length 2682;
Best Local Similarity 42.1%; Pred. No. 0.00061;
Matches 453; Conservative 0; Mismatches 624; Indels 0; Gaps 0;

Qy 324 CGACACCCCCCGAGCGAGCGCGCGCGAGCGGACCCCTGACTTCCCGCAGAT 383
Db 501 CTATCTGAGAGCATCTGCTCGAGGGGTTCCCGACCGCGCGGCGCACTTCACCTGCAA 560

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 1, 2005, 07:40:05 ; Search time 7875.14 Seconds  
(without alignments)  
11933.846 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469  
Sequence: 1 gtgagccaccatgcccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.\*

1: gb\_est1.\*  
2: gb\_est2.\*  
3: gb\_hic.\*  
4: gb\_est3.\*  
5: gb\_est4.\*  
6: gb\_est5.\*  
7: gb\_est6.\*  
8: gb\_gss1.\*  
9: gb\_gss2.\*

\* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	136.4	5.5	330	9	CL293849
2	97.8	4.0	2031	9	CL974989
3	97.2	3.9	2886	9	CL967755
4	91.4	3.7	1509	9	CL959255
5	90.8	3.7	2598	3	AY103647
6	86.6	3.5	1398	9	CL961989
7	85.6	3.5	951	4	BM321451
8	85	3.4	869	7	CK159167
9	84	3.4	3069	9	CL973991
10	83.8	3.4	1941	9	CL971508
11	80.2	3.2	1132	4	BM320864
12	79.6	3.2	1165	4	BM320900
13	79	3.2	867	4	BM321430
14	78.4	3.2	11691	9	CL962901
15	78.2	3.2	1962	9	CL961326
16	78.2	3.2	2853	9	CL974397
17	78	3.2	1485	9	CL970981
18	76.6	3.1	1550	4	BM321022
19	76.6	3.1	2559	9	CL982027
20	76.2	3.1	1401	9	CL962721
21	75.6	3.1	671	6	CA093222
22	75.4	3.1	2299	3	AY106831
23	75.2	3.0	854	7	CK777127
24	75	3.0	862	4	BM321023

## RESULT 1

CL293849/c

LOCUS

DEFINITION

02S0349-08A1-C03 UniformMu MUTAIL Library Zea mays genomic clone

02S0349-08A1-C03, genomic survey sequence.

ACCESSION

CL293849

VERSION

CL293849.1

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Zea mays

REFERENCE

1 (bases 1 to 330)

AUTHORS

Latschew S., Tan, B.-C., Settles A.M. and McCarty, D.R.

TITLE

Sequence tagged transposon insertions from the UniformMu maize population

JOURNAL

Unpublished (2003)

COMMENT

Contact: Donald R. McCarty

Plant Molecular and Cellular Biology Program

University of Florida

PO 110690 Gainesville, FL 32611-0690, USA

Tel: 352-392-1928 x322

Email: drmc@ufl.edu

Sequence flanking probable Mu insertion site in UniformMu line:

02S0349-08, Primer set: A

Class: transposon insertion site.

Location/Qualifiers

1. .330

source

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="W22 (ACR, bz1-m9)"

/cultivar="UniformMu"

/db\_xref="taxon:4577"

/clone="02S0349-08A1-C03"

/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon

insertions in Mu inactive lines were extracted from the

UniformMu maize population by the thermo asymmetric

interlaced PCR (TAIL) protocol using primers specific for

the Mu terminal inverted repeat and a set of 16 arbitrary

primers. Amplicons were size enriched using Sepharose 400

spin columns and cloned into the TOPO PCR4 vector."

## ORIGIN

Query Match 5.5%; Score 136.4; DB 9; Length 330;  
Best Local Similarity 66.4%; Pred. No. 2.3e-14;

## ALIGNMENTS

CL293849 330 bp DNA linear GSS 12-FEB-2004  
02S0349-08A1-C03 UniformMu MUTAIL Library Zea mays genomic clone  
02S0349-08A1-C03, genomic survey sequence.  
CL293849  
CL293849.1 GI:42541978  
GSS.  
Zea mays  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 330)  
Latschew S., Tan, B.-C., Settles A.M. and McCarty, D.R.  
Sequence tagged transposon insertions from the UniformMu maize  
population  
Unpublished (2003)  
Contact: Donald R. McCarty  
Plant Molecular and Cellular Biology Program  
University of Florida  
PO 110690 Gainesville, FL 32611-0690, USA  
Tel: 352-392-1928 x322  
Email: drmc@ufl.edu  
Sequence flanking probable Mu insertion site in UniformMu line:  
02S0349-08, Primer set: A  
Class: transposon insertion site.  
Location/Qualifiers  
1. .330  
source  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="W22 (ACR, bz1-m9)"  
/cultivar="UniformMu"  
/db\_xref="taxon:4577"  
/clone="02S0349-08A1-C03"  
/notes="Vector: TOPO-PCR4; DNA flanking Mu transposon  
insertions in Mu inactive lines were extracted from the  
UniformMu maize population by the thermo asymmetric  
interlaced PCR (TAIL) protocol using primers specific for  
the Mu terminal inverted repeat and a set of 16 arbitrary  
primers. Amplicons were size enriched using Sepharose 400  
spin columns and cloned into the TOPO PCR4 vector."

25 75 3.0 3822 9 CL972913  
26 74.6 3.0 914 8 BZ568300  
27 74.6 3.0 2010 9 CL962831  
28 74.6 3.0 2313 9 CL982362  
29 74.6 3.0 2682 9 CL969033  
30 74.4 3.0 3134 3 AY109500  
31 74.2 3.0 1680 9 CL982770  
32 74 3.0 1290 9 CL972679  
33 74 3.0 2072 3 CR603312  
34 73.8 3.0 889 7 CK159613  
35 73.6 3.0 757 9 CC678788  
36 73.6 3.0 2028 9 CL979437  
37 73.2 3.0 853 4 BM321393  
38 73 3.0 1689 9 CL972373  
39 72.8 2.9 566 4 BM587428  
40 72.8 2.9 1575 9 CL979927  
41 72.6 2.9 788 6 CB643171  
42 72.4 2.9 753 9 CC675888  
43 72.4 2.9 801 4 BJ375394  
44 72.2 2.9 688 6 CB648640  
45 72.2 2.9 764 6 CB651670

Matches	211;	Conservative	0;	Mismatches	106;	Indels	1;	Gaps	1;	
Qy	421	GGCGGGCCAGATCAAGGAGGCGCTGTGTGGACACCGCGCGCGACGACACCGTGTCTGGAGGAG	480							
Db	329	GGGGGGCAGCTGGAAGAAGCTCTATTAGATACAGGAGCAGATGATACAGTATTAGAAGAA	270							
Qy	481	ATGAGCGTCCCGGCAAGTGGGAAGCCCAAGATGATCGCGGGCATCGCGGGCTTCAATCAAG	540							
Db	269	ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGAAATTGGAGGTTTTTATCAAA	210							
Qy	541	GTGCGCAGTACGACAGATCTGTGTCGAGATCTGGGCNAGAAGGCCATCGGCACCGTG	600							
Db	209	GTAAGACAGTATGATCAGGTACCCATAGAAATCTGTGGGCATATAAGCTATAGGTACGGTA	150							
Qy	601	CTGATCGGCCCCACCCCGTGAACATCATCGCGCGCAACATCTCACCCAGCTCGGGCTGC	660							
Db	149	TTAGTAGGACCTACACTGTCAACATAATTTGAAGAAATCTGTTGACTCAAGTTGG-TGC	91							
Qy	661	ACCTGAACTTCCCCATGACCCCCATCGAGACCGTGCCTCGTGAAGCTGAAAGCCCGGCATG	720							
Db	90	ACCTTAAATTTTCCCATTAGTCTTATGAACTGTGACCAGTAAATTTAAAGCCAGGAATG	31							
Qy	721	GACGGCCCCAGGTGAAG	738							
Db	30	GATGGCCCCAAAAGTAAGG	13							
RESULT 2										
LOCUS	CL974989	2031 bp	DNA	linear	GSS	21-SEP-2004				
DEFINITION	OsIFC042895 Oryza sativa Exress Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.									
ACCESSION	CL974989									
VERSION	CL974989.1	GI:52404497								
KEYWORDS	GSS.									
SOURCE	Oryza sativa (indica cultivar-group)									
ORGANISM	Oryza sativa (indica cultivar-group)									
REFERENCE										
AUTHORS	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G.K.S., Deng, X.W. and Wang, J.									
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis									
JOURNAL	Unpublished (2004)									
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.									
FEATURES										
source	Location/Qualifiers									
	1..2031									
	/organism="Oryza sativa (indica cultivar-group)"									
	/mol_type="genomic DNA"									
	/db_xref="taxon:39946"									
	/clone_lib="Oryza sativa Exress Library"									
	/note="Oryza sativa exon trapped genomic sequences "									
ORIGIN										
	Query Match	4.0%;	Score	97.8;	DB	9;	Length	2031;		
	Best Local Similarity	42.6%;	Pred.	No. 2.2e-07;						
	Matches	787;	Conservative	0;	Mismatches	1032;	Indels	30;	Gaps	4;
Qy	576	CGGCAAGAAGGCCATCGGCACCGTGTGATCGGCCCGCCCGGTGAACATCATCGGCCG	635							
Db	153	CGGCACGACGACTACTCGTGGCTGGAGTCTTACCGGAACCGCCATCGAGATCATCGCAA	212							

```
Db 1281 GAAACAGGAGAGCATGATCTCTCGACGTGCGCGCGTCAACCTTCGGCTTGGAGACGGC 1340
Qy 1773 GCTGACGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTGATCTGGGCAAGACCCC 1832
Db 1341 CGCGCGGTGATGCGCAAGCTGATCCCGCGCAACACGCTGCTGCGGAGAGACGCA 1400
Qy 1833 CAAGTTCCGCTGCCATCGAAGAGAGACCTGGGAGACCTGGTGGACCGACTACTGGCA 1892
Db 1401 GGTGTTTACCACGCTACAGGAGCAAGACGACAACTGACCATCCAGGTGTTTCGAGGGCGA 1460
Qy 1893 GGCACCTGGATCCCGAGTGGAGTTCGTGAACACACCCCGCTGCTGGAAGCTGTGGTA 1952
Db 1461 GCGGAGCATGACGAGGAGCAACCGGCTGCTCGGAGGTTCGACCTTCGCGGCAATCGCGCC 1520
Qy 1953 CCAGCTGAGAGAGAGCGCCATCATCTGCGCGCGAGACCTTCTACGTGAGCGCGCGCCAA 2012
Db 1521 GCGCGGAGGGGCGCGCAGATCGAGTGAAG-----TTCGAGGTGGAGCGAA 1571
Qy 2013 CCGGAGACCAAGATCGGCAAGCGCGTCTAGTGAACCGACCGGGCGCGGCAAGATCGT 2072
Db 1572 CGGCATCTCAGCGTGTGCGCGCGCAAGAGCCACCGGAGGTTCGAGAGATCACCAT 1631
Qy 2073 GAGCTGACGAGACCAACCAAGAGACCGAGTGCAGGCTCCAGCTGGCGCTTGA 2132
Db 1632 CTCGCGGACACCGCAAGATCAGCCAGGAGATCGACCGGATGTGCGCGAGCGGA 1691
Qy 2133 GGACAGCGGAGAGCGAGTGAACATCTGACCGACAGCGAGTACCGCTTGGGCTATCTCCA 2192
Db 1692 GGAGTTCCGCGAGGAGACCGCGGACAGGAGCGGTGACCGCGGAGACGCTTGA 1751
Qy 2193 GGCCAGCGGACCAAGAGCGAGAGCGAGTGTGTGAACAGAT-----CATCGAGAGCT 2246
Db 1752 GCGCTAGCTCTACAACTCAAGAACAGCTCGCGGCAAGATGCGGAGCGCATGTGAGGG 1811
Qy 2247 GATCAAGAGAGAGAGTGTACCTGAGTGGTGGTCCCGCCACAGGCGCATCGCGGCA 2306
Db 1812 CGAGGAGAGAGAGAGTGTGAGGAGCGGTGAGGAGCGGTACAGTGGTGGACGCGAA 1871
Qy 2307 CGAGCAGATCGCAAGCTGTGTGAGCAAGGGCATCCGCAAGTGTGTCTTGGACGGCAT 2366
Db 1872 CCGGAGCGCGCAAGGAGGAGTACGAGGAGAGCTGAGGAGTGTGAGGAGCTGTGCA 1931
Qy 2367 CGATGGCGGCATCGTATCTACAGTATACGAGCATCGACCTGTACGTGGC 2415
Db 1932 CCGCGTATGTCGCGCGTCTACAGAGAGTCCGCGGCGCGCGCGCGC 1980
```

## RESULT 3

```
CL967755
LOCUS CL967755
DEFINITION OsiFCC015718 Oryza sativa Expressed Sequence Tag (EST) Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL967755
VERSION CL967755.1 GI:52390149
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 2886)
Ma, L., Wang, J., Chen, C., Liu, X., Sun, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G. K. S., Deng, X. W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
```

```
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source 1..2886
Location/Qualifiers
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Expressed Library"
/note="Oryza sativa exon trapped genomic sequences"
```

## ORIGIN

```
Query Match 3.9%; Score 97.2; DB 9; Length 2886;
Best Local Similarity 43.2%; Pred. No. 2.8e-07;
Matches 712; Conservative 0; Mismatches 903; Indels 33; Gaps 4;
Qy 189 GATGAAGAGCTGACACGAGCGCCAGGCACTTCTTCGGGAGGAGACCTTGGCCCTTCCCA 248
Db 219 GGTGATGAGCGCGCCGCCATCATGGCCATCTCTCGCCAAACGGCGCGGCGGCGCC 278
Qy 249 GGGCAAGCGCGCGAGTTCCTCCAGCGAGCAAAACCGCGCAACAGCCGCCACAGCGCGCGA 308
Db 279 CGACTGAGAGACTTCTGCGCATCTGCTCTCTCTCATCACTCCACCATCAGTTT 338
Qy 309 GTGCAAGTGGC---GGCGACAAACCCCGAGAGCGCGCGCGCGCGCGCGCGCGCGAC 365
Db 339 CATCGAGGAGAAACAATCCCGGAAACCGCGCGCGCTCATGGCGCGCTTCGCGCCAA 398
Qy 366 CTTGAATCTCCCGAGATCACTGTCGAGCGCGCGCTCTGTCGATCAAGTGGCGG 425
Db 399 GACCAAGTGTCTAGGAGCGGGAATGGAGGAGAGAGAGCGCTTCCTCTGTTCCCG 458
Qy 426 CCAGATCAAGAGAGCGCTGTGGAACCGCGCGCGAGCACACCGCTGTGAGGAGATGAG 485
Db 459 CGACATCATCAGCATCAAGCTCGCGACATCATCCCGCGAGCGCGCGCTGCTCGAGG 518
Qy 486 CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGG 545
Db 519 CGACCGCTCAAGTGTGACAGCGCGCGCTCACCGCGAGTTCGATGCGCTCAACAAGCA 578
Qy 546 CCAGTACAGCAGATCTGATCGAGATCTGCGGCAAGAGCCATCGGACCGTCTGAT 605
Db 579 CGCGGCGAGGGCGTCTTTCGCGGTCCACCGTCAAGAGCGGCGAGATCGAGGCGCTG 638
Qy 606 CGGCGCGACCGCGTGAACATCATCGCGCGCAACATGTCACCGAGCTGGGCTGCACCT 665
Db 639 CATGCGCACCGCGGTGACACCTTCTTCGCGAGCGCGCGCACCTGGTGGACAGCA 698
Qy 666 GAATCTCCCATCAGCGCCATCGAGACCGTTCGCGTGAAGCTGAAGCGCGGATGGAG 725
Db 699 CAACATCGGCGCACTTCAGCTGCTGCTCAGCGCATCGGCAACTTCTGTCATCATCTCC 758
Qy 726 CCCCAGGTGAAGAGTGGCGCTGACCGAGGAGAGATCAAGGCGCTGACCGCGCATCTG 785
Db 759 CGCGTCCGATGATATCGAGATCATCTGTTATGATACCGCATCCAGACCGCGGCTAC 818
Qy 786 CGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGGCGCGCGAGAACCCCTACAA 845
Db 819 CGAGCGATGACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 878
Qy 846 CCGCGTGTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 905
Db 879 CGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 938
Qy 906 CGAGTGAACAGGCGACCGGAGCTTCTGGAGGTGAGTGGGATCCCGCGCGCGCGCG 965
Db 939 CAAGCGGATGACCGCGCATCGAGGAGATGGCGCGCATGAGCGTCTCTGTCAGCGCA 998
Qy 966 CGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025
Db 999 CGGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1058
```



```

Qy 1150 TTCAGAGCAGCATGACCAAGATCTTGAGCCCTTCGCGCCGCCCAACCCCGAGATCGTG 1209
Db 721 AGCTACAAACAGCAGCCCAACAGCTCGCGTCCGACGCGCAGCTCGGTGCTCTTC 780
Qy 1210 ATCTACCAAGTACATGAGGAGCTGTAGTGGGAGAGACCTGGAGATCGGCCACACCGC 1259
Db 781 TGGATGTCCATCGCGGAGGAGCTCGCGTGTGCTGCGCCCGCGCTCGAGCGGTATC 840
Qy 1270 GCCAAGATCGAGGAGCTGCGCAAGCACCTGTGCGCTGGGCTTTCACACCCCGACAAG 1329
Db 841 CTGTGGAACGGGTGCCCTTGACCCGCGGTGTCGCGCTGATTTCGACACAGGTCAG 900
Qy 1330 A 1330
Db 901 A 901

RESULT 5
AY103647
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2598)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 2598)
Coe,E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
Location/Qualifiers
1..2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

FEATURES
source
Query Match 3.7%; Score 90.8; DB 3; Length 2598;
Best Local Similarity 43.3%; Pred. No. 4e-06;
Matches 719; Conservative 0; Mismatches 909; Indels 33; Gaps 5;

Qy 170 GCGCAAGAGGGGCCACAGATGAAGGACTGCACCGAGCGCGCAGGCCAACATCTTCGCG 229
Db 351 GCACGAGTGGATCTCTGACTGCTGCGCCACGCGGGGCGACATCTACGGGTGACACCG 410
Qy 230 AGGACCTGGCTTCCCGGAGGCGCGCGAGTTCGCCAGCGAGCAGAACCGGCCA 289

```

```

Db 411 GCTTCGCGGCACCTCCACCGCGCACCAAGGACGGGCCCGCTCCAGGTCCAGGTGC 470
Qy 290 ACAGCCCAACAGCGCGGAGCTGCAGGTGCGCGCGCAACCCCGCAGCGCGCGCG 349
Db 471 TCAGGCATCTCAACCGCGGAATCTTCGGCACCGGCGGACACAGCTCGCGTGG 530
Qy 350 CCGAGCGCGAGGGCAACCTTGAACTTCCCGCAGATCACCTGTGGAGCGCCCCCTGTGTA 409
Db 531 AGGTCAACCGCGCGCGATGCTGTGGCATCAACACCCCTCTCCAGGGTACTCCGCA 590
Qy 410 GCATCAAGGTGGCGGCGCAGATCAAGGAGGCCCTGTGACACCGCGCGCGACGACCG 469
Db 591 TCCGCTTCGAGATCCTCGAGGCATCAGAACTGCTCAACCGGTGTACGCCCTGCC 650
Qy 470 TGCTGGAGGAGATGAGCTTCGCCGCAAGTGGAAAGCCCAAGATGATCGCGGCATCGCG 529
Db 651 TGCGCTCCGGGCAACCATCACCGGTGCGGCGAGCTGGTCCGCTCTCTTACATCGCG 710
Qy 530 GCTTCATCAAGGTGCGCGCATGACACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCA 589
Db 711 GCCTCATCAGCGCGCGCCCAACCGCGAGGCGCTCACCG---TCGACGCGAGGAGGTGG 767
Qy 590 TCGCACCGTCTGATCGCGCCCAACCCCGTGAAATCATCATCGCGCGCAACATGCTGACCC 649
Db 768 ACGCCCGCGAGCGGCTTCAAGATCGCGGCATCGAGGGCGGCTTCTTCAAGCTTCAACCCCA 827
Qy 650 AGCTGGGCTGCACCCCTCAACTTCCCATCAGCCCCCATCGAGACCGTGCCTGTAAGCTGA 709
Db 828 AGGAGGCTCTGCATCTGTCAACCGGCGCTCGGTGGGCTCGCGCTCGCGGCCACCGTGA 887
Qy 710 AGCCCGCATGAGCGGCGCCCAAGGTGAAGTAGTGGCCCTGACCGAGAGAGATCAAGG 769
Db 888 TGTAACGACGCAACGCTCTGCGCGCTCTGTGCGAGGTCTCTGTCGCGCTCTTCTGCGAGG 947
Qy 770 CCCTGACCGCATCTGCGAGGAGATGAGAGAGGAGGCGCAGATCACCAAGATCGGCCCG 829
Db 948 TCATGAACGCGCAGCCCGAGTACACGACCACTGACCCCAACAGTGAAGCACCACCGG 1007
Qy 830 AGAACCCCTCAACACACCCCGTGTTCGCTCAAGAGAGAGAGGACAGACCAAGTGGCGCA 889
Db 1008 GGTCCATCGAGGCGCGGCCCATCATGAGGCACATCTTGGATGGAGCTCTTTCATGAAGC 1067
Qy 890 AGCTGTGGACTTCCGCGAGCTGAACAGGCGCACCCAGGACTTTGGAGAGTGCAGTGG 949
Db 1068 AGGCAAGAAGTTGAACGAGCTGACCCGCTGAGACCCGCTGCTGAGAGGAGTCTTCTGCGAGG 1115
Qy 950 GCATCCCGCCCGCGCGCTGAAAGAGAGAGAGGCTGACCGTCTGCGAGTGGGCG 1009
Db 1116 ACAGGTACGCGCTCCGACAGTCCGCGAGTGGTGGGCGCCCGCAGATCGAGGTCTATCCGCG 1175
Qy 1010 ACGCTACTTTCAGCGTGCCCTCGAGCAGGAGCTTCCGCAAGTACACCGCTTTCACCATCC 1069
Db 1176 CCGCCACCAAGTCCATCGAGGCGAGGTCACTCGGTGAGACGACACACCGGTCTATCGAGC 1235
Qy 1070 CCAGCATCAACACGAGACCCCGGCATTCGCTACCAAGTACAAAGTGTGTGCGCCAGGGCT 1129
Db 1236 TCCACCGCGGCAAGCGCTGCACGCGGCAACTTCCAGGGCACCCCGCATCGCGGTGTCCA 1295
Qy 1130 GGAGGGCAGCCCGAGCATCTTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCGCG 1189
Db 1296 TGGCAACGCGCGCTCGCCATCGCCCAACATCGGCAAGTCTATGTTGCGCAGTCTCTCCG 1355
Qy 1190 CCGCAACCCCGAGATCTGATCTACCAGTACATGGACGAGCTGTAGCTGGGCGAGCGCA-- 1247
Db 1356 AGCTGTCAAGAGATTCTACAAAGCGGCTCACTCCACCTGGCGGAGCGCGCAACC 1415
Qy 1248 ----CCTGGAGATCGGCCAGCACCGCGCAAGATCGAGGAGCTCGCGAAGCACCTGTGTC 1303
Db 1416 CCAGCCTGGACTACGGCTTCAAGGGCACCGAGATCGGCATGCGCTCTCTACTGCTCCGAGC 1475
Qy 1304 GCTGGGGCTTCAACACCCCGCAAGAGAGCAGACCAAGAGAGGAGCCCGCTTCTCTGTGGATGG 1363

```

```
Db 1476 TCAGTACCTGGGCAACCCCATCACCAACCACTGCAGAGCGCGGAGCAGCACAAACGAG 1535
Qy 1364 GCTACGAGCTGCACCCCGGACAACTGACCGTGCAGCCCATCGAGCTGCCGCGAAGGAGA 1423
Db 1536 ACGTGAATCCCTGGGCTCTGCTCGGCGCAGGAAGACCGCGAGGCGATCGACATCCTGA 1595
Qy 1424 GCTGGACCGTGAACGACATCCAGAACTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT 1483
Db 1596 AGTCTATGTCGTCACTACATCGTGGCGTGTGCCAGGCGGTGGACCTGGCGCACTCG 1655
Qy 1484 ACCCGGCAATCAAGGTGCGCCAGCTGTGCAAGTGTGTGCGCGCGCAAGGCCCTGACCG 1543
Db 1656 AGGAGAACTCAAGGGCTGTGTTGAAGAACACCGCTGACCCAGCTGGCGGCAAGAGGTGTA 1715
Qy 1544 ACATCTGCCCCCTGACCGAGGAGCGAGCTGAGCTGGCCGAGAA-----CGCG 1594
Db 1716 CCATGAACCCCTCGGGCGAGCTCTCCAGCGCCGCTTTCAGCGAGAAAGGAGCTGATCAGCG 1775
Qy 1595 AGATCTGCGCGAGCCGTCGACGGGTGTACTACGACCCCAAGGACCTGGTGGCG 1654
Db 1776 CCATGACCCGAGGCGGTGTTACGTACCGGAGGACGCGGCGCAGCGAGCTGCGCG 1835
Qy 1655 AGATCAGAGCAGGGCCACGACGAGTGGACCTTACCAGATCTACGAGGAGCCCTTCAAGA 1714
Db 1836 TGATGAGAGCTGCGCGCGTCTGTGTGACACGCGCTCAGCAGCG---GGACGCG 1892
Qy 1715 ACCTGAGACCGCAAGTACGCAAGATGGCGACCGCCACACCAAGGAGTGAAGCAGC 1774
Db 1893 AGCGGAGGCGCTCCGTGTTCTCAAGATCACAGGTTTCGAGGAGGAGCTCCGCGCGGTG 1952
Qy 1775 TGACCGAGGCGGTGACAGAGATCGCATGGAGAGCATCGTG 1815
Db 1953 TGCCCCAGGAGGTGGAGGCGCGCGGTGCGGTGCGCGAG 1993
```

## RESULT 6

```
CL961989
LOCUS 1398 bp DNA linear GSS 21-SEP-2004
DEFINITION OIIFC006991 Oryza sativa Express Library Oryza sativa (indica
cultivar-group) genomic, genomic survey sequence.
ACCESSION CL961989
VERSION CL961989.1 GI:52378720
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1398)
AUTHORS Ma,L., Wang,C., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source
1..1398
Location/Qualifiers
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Exprim Library"
/note="Oryza sativa exon trapped genomic sequences "
```

## ORIGIN

```
Query Match 3.5%; Score 86.6; DB 9; Length 1398;
Best Local Similarity 45.9%; Pred. No. 2.2e-05;
Matches 508; Conservative 0; Mismatches 569; Indels 30; Gaps 5;
Qy 1 GTCGAGCCCAACATGCGCCGAGCCATGAGCCAGGCCACAGGCCCAACATCTCTGTATGTCAG 60
Db 166 GTCGAGAGCAGCTGCGCGAGGCGGACATCTTCTGCTCCGTGAACACCCCAACCAAG 225
Qy 61 CGCAGCAACTTCAAGGGCCCCAAGCCCATCATCAAGTGTCTTCACTGGCGCAAGGAGGGC 120
Db 226 GCCCGCGGCTCTGCGCCCGCAAGCGCCG---ACCTCACCTACTGGGAGAGCGCGCG 282
Qy 121 CACATCGCCGCAACTGCGCGCGCCCCCGCAAGAGGCTGTGTGAAGTGCGGCAAGGAG 180
Db 283 CGGATGATCGCGCGCTGGCGAGCTCGGACAAGTCTGTCGAGAAGTCCACCGTGCCG 342
Qy 181 GGCACACAGATGAAGACTGACCCGAGCGCGAGGCCAACTTCTTCGCGAGGAGCCTTGCC 240
Db 343 GTCAGAGCGCGGAGGCCATCGAGAAGATCTTCGACCAACAGCGCGCGAGCGCGTGGC 402
Qy 241 TTCCCCCGAGGCAAGGCCCGCGAGTTCCCGAGCGAGCAGAACCGCGCAACAGCCCCACC 300
Db 403 TTCAGATCTCTTCAACCCGAGTTCTCTGCGGAGGACACCCCATTCGCGGACCTGTCT 462
Qy 301 AGCGCGAGCTGCAGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGAGCGCCAG 360
Db 463 GCCCCCGA-----CGCGTCTCATCGGCGCGCGGAGACCGCGCGGCGCGCC 513
Qy 361 GGCACCTGAACTTCCCCAGATCACCTGTGTGCGAGCGGCCCTCTGTGAGCATCAAGGTG 420
Db 514 GCGTCCAGCGCTCAAGGACGTGTACGCGCGATGGGTCCCCGAGGAGGATC---CTC 570
Qy 421 GCGGCGAGATCAAGGAGGCCCTGCTGGACACCGCGCGCGACACCGCTGTGGAGGAG 480
Db 571 ACCACCAACTCTGTGTCGCGGAGCTGTCAAGCTCGCGCCAAACGGTCTCTTGCGCCAG 630
Qy 481 ATGAGCTTCGCGGCAAGTGGAGCCCAAGATGATCGCGGCGCATCGCGCTTTCATCAAG 540
Db 631 AGGATCTGCTGCTCAACGCCATGTCCGCGCTCTGCGAGGCCACCGCGCGCGAGCTCGCC 690
Qy 541 GTGCGCAGTACGACAGATCTCTGATCGA---GATCTGGGCGAAGAGGCCATCGGCACC 597
Db 691 GAGGTGCGCTACGCGCTGGGCAAGGACTCCAGGATCGGCGCCAAAGTTTCTCAACGCGCAG 750
Qy 598 GTCTGATCGGCGCCACCCCGGTGAACATCATCGCGCGAAACATGCTGACCCAGCTGGGC 657
Db 751 GTGGGATTGCTGGCTCTGCTGCTTCCAGAAGACATCTCAACCTGTG-----CTG 798
Qy 658 TGCACCTTGAACCTTCCCATCAGCCCCATCGAGACCGTGCCTGAGCTGAAGCCCGCGC 717
Db 799 TACATCTGCGAGTGCAATGGCTCTCCCGAGGTGGCCAACTACTTGAAGCAGGTTATCAAG 858
Qy 718 ATGAGCGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGGCCCTTGACC 777
Db 859 ATCAACGACTTACCAGAAGAGCCGTTCTGTAACCGGTGTGCTGCTCATTTCAACACC 918
Qy 778 GCCATCTGCGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGSCCCCGAGAACCCC 837
Db 919 GTGGCGGCAAGAAAGATCGCGGTCTGGGTTCTGCGCTTCAAGAGGACACCGCGGACACC 978
Qy 838 TACAACACCCCCGTGTTTCGCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAGCTGGTG 897
Db 979 AGGGAGACGCCGCCATCGACGTCTCAAGGGGCTGATCGGCGCAAGGCCAAGGTGAGC 1038
Qy 898 GACTTCGGGAGTGAACAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCGATCCCC 957
Db 1039 ATCTAGACCCCCCAGGTGACGGAGGCCAGGTCTCAGCGGACCTCGCCATGAGCAAGTTC 1098
Qy 958 CACCCCGCGGCTGAAGAAGAAGAGAGCGTCAACCGTGTGACCGTGGCGCGACCGCTAC 1017
Db 1099 GATTGGATCACCCGTTCCATCTCCAGCGGATGAGCCCCACGCCCATCAAGAGGAGTTC 1158
Qy 1018 TTCAGCGTGCCTGGACGAGGACTTCCGCAAGTACACCGCGCTTACCATCCCCAGCATC 1077
```

```

Db 1159 GTCCCTTGGGACGCTTACGAGCGCGCCAGGACGCCACCGCGTGTGATCCTTCAACGAG 1218
Qy 1078 AACACGAGACCCCGGCGATCCGCTAC 1104
Db 1219 TGGGACGAGTTTCAGGAGCGCTGACTAC 1245

RESULT 7
BM321451 951 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller 0.1211 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to adenosylhomocysteinase (EC
3.3.1.1), mRNA sequence.
ACCESSION BM321451
VERSION 1
KEYWORDS EST.
SOURCE BM321451.1 GI:18055857
ORGANISM Mastigamoeba balamuthi
Eukaryote; Pelobiontida; Mastigamoebidae; Mastigamoeba.
REFERENCE 1 (bases 1 to 951)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 951 Std Error: 0.00
POLYA-No. Location/Qualifiers
FEATURES
source 1..951
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/notes="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match 3.5%; Score 85.6; DB 4; Length 951;
Best Local Similarity 46.0%; Pred. No. 3.2e-05;
Matches 323; Conservative 0; Mismatches 375; Indels 3; Gaps 1;

Qy 1734 CGCCAGATCGGCACCGCCGCACACACGAGCGTGAAGCAGTGAACCGGCGGTGCAGAA 1793
Db 219 CGCCAGATCGGTGGTGTCTGTCGACATCTTCTCGACGAGGACACCGGCGCGCG 278
Qy 1794 GATCGCATGAGAGCATGTGATCTGGGCAAGACCCCAAGTTCGCTGCCATCCA 1853
Db 279 CATCGCAGCGCGCGTCTCGGTCTTCGCTGGAAGGCGGAGAACCTCCAGGAGTACTG 338
Qy 1854 GAAGGACCTGGGAGACCTGTGGACGACTATCGGCGCCACCTGGATCCCGAGTG 1913
Db 339 GGAGTGCACTGGAAGGCGCTGTCTTCGCGCCCTACGAGGCGCTCAGATCATCGTCGA 398
Qy 1914 GGAGTTCTGTAACACCCCGCCCTTGGTGAAGCTGTGTACGCTGGAAGAGGAGCCCAT 1973
Db 399 CGACGGCGGTGACCGACTCTGATGATCCACNAGGGGTTTCGGCGCGGACACCCCAA 458
Qy 1974 CATCGGCGCGGAGACCTTCTACGTGGAAGCGCGCGCCCAACCGGAGACCAAGATCGCAA 2033
Db 459 GCTGCTGGAGGACGACGAGGCGCTTCGAGGAGGTTCGCTGCTCAACACGCTGCTCAAGCA 518
Qy 2034 GCGCGGTCTAGTACCACCGGCGCGCGGAGAGATCGTGTGCTGACCGGAGACCCAA 2093
Db 519 GGTCCAGAGAGGACGCGCGGCTTCTGGCAAGATCCTCCCCGAGATCCCGCGGTGTCAG 578

```

```

Qy 2094 CCAGAAGACCGAGTGTGAGGCGCATTCAGTGGCCCTTCAGACACGCGCAGCGAGGTCAA 2153
Db 579 CGAGGAGACGACGACTGGCGTGTATGAGGCTGTACAGCTGCACGCGCAGCGGAGCTGCT 638
Qy 2154 CATGCTGACCGACAGCCAGTACGCGCTGGGCGATCATCCAGGCCCGGCGGAGGAGCA 2213
Db 639 GTTCCCGCGCGTCAACGTCACGAC---TCTNTTCAACAAGAGCAAGTTTCGACAACATCTA 695
Qy 2214 GAGCGAGTGTGTAAACAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTCGAG 2273
Db 696 CGGTTCGCGCACTCGCTCATCGAGCGCATCAAGCGGCGGACCCACGTCGTCTCGGCGG 755
Qy 2274 CTGGGTCCCGCCCAAGGCGCATTCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAA 2333
Db 756 CAAGGTGCGCGTCTGTCGCGGCTACGCGGAGCTGGGCAAGGCGTGGCGGAGTGGCTCGG 815
Qy 2334 GGGCATCCGCAAGGTGTCTTCGAGCGGCGCATCGATGGCGGCGATCGTGATCTACCAGTA 2393
Db 816 CGGCGAGGCTGCGCGCTCATCTGTGACGAGATCGACCCCATCTGCGCGCTGCAGGGGTC 875
Qy 2394 CATGAGCAGCTGTACCTGCGGCGGCGGCGGCGCTAGGATCGA 2435
Db 876 GATGGCGGCTTCGAGGTCAACGCTCGAGGCGGCGCTCGA 917

```

```

RESULT 8
CK159167/c 869 bp mRNA linear EST 05-DEC-2003
LOCUS FGAS040564 Triticum aestivum FGAS: TaL5 Triticum aestivum cDNA,
DEFINITION mRNA sequence.
ACCESSION CK159167
VERSION CK159167.1 GI:38985053
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum

```

```

REFERENCE 1 (bases 1 to 869)
AUTHORS Allard,P., Crosby,M.L., Danyluk,J., Eudes,P., Frick,M., Gaudet,D.,
Genswein,B., Graf,R., Gulick,P., Hrycan,L.D., Laroche,A.,
Links,M.G., McCarthy,E.L., Monroy,A., Muzak,I., Nilson,D.,
Penniket,C., Roach,J.L. and Sarhan,F.
TITLE Functional Genomics of Abiotic Stress In Wheat and Canola Crops
JOURNAL Unpublished (2003)
COMMENT Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas_este@cs.usask.ca

```

```

This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [128,636].
Plate: TaL537 row: N column: 23.

```

```

FEATURES
source Location/Qualifiers
1..869
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: TaL5"
/notes="Organ: Crown; Vector: pGEM-T; SSH (suppression
subtractive hybridization) cDNA library from genotype
PI178383 cold hardened at 2 C for 21 days and 49 days
(equal amount of cDNA pooled together before subtraction,
tester) and subtracted against genotype Norstar cold

```



hardened at 2 C for 1 day (24 H) (driver). Modified Smart  
cDNA (Clontech) priming and non-directional cloning"

ORIGIN		Query Match	3.4%;	Score 85;	DB 7;	Length 869;	
		Best Local Similarity	45.4%;	Pred. No. 4.1e-05;			
		Matches 304;	Conservative	0;	Mismatches 365;	Indels	0; Gaps 0;
QY	580	AAGAAGCCATCGGACCGTGTGATCGGCGCCACCCCGTGAACATCATCGCGCCGCAAC	639				
DB	827	ATGGCGCGCGCCGCAACAGAGGATCACCAACAGGACGAGAACCAACCAACACCAAC	768				
QY	640	ATGCTGACCCAGCTGGCTGCACCTGAATCTCCCATCAGCCCTCATCGAGCCGTCGCC	699				
DB	767	CACACGACGACACAGGACACACGACCAACCAACACACACACACACACACAC	708				
QY	700	GTGAAGCTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAG	759				
DB	707	ACCAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	648				
QY	760	AGATCAAGCCCTGACCGCTCTGGAGGAGATGGAGAGGAGGCGGACGATCACCAAG	819				
DB	647	AACAACACGACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	588				
QY	820	ATCGGCCCGGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAGAGGACGACACC	879				
DB	587	AACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	528				
QY	880	AAGTGGCGCAAGTGTGGACTTTCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGGAG	939				
DB	527	AACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	468				
QY	940	GTGAGCTGGGATTCCTCCACCCCGCGGCTCAAGAGAGAGAGCGTGCCTGCTG	999				
DB	467	AACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	408				
QY	1000	GAGTGGGCGACGCTTACTTTCAGCTGCCCTCGAGGAGACTTCGCGAAGTACACCGCC	1059				
DB	407	GACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	348				
QY	1060	TTCAACATCCCGAGCTCAACACGAGACCCCGGATCCGCTTACGATCAACAGTGTG	1119				
DB	347	GACCAACAAACGACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	288				
QY	1120	CCCGAGGCTGGAAGGCGACCCCGAGTCTTCAGAGGAGATGACCAAGATCTCTGGAG	1179				
DB	287	AACAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC	228				
QY	1180	CCCTTCGCGCCCGCAACCCCGAGATCGTGATCTACAGTACATGACGACCTGTACGTG	1239				
DB	227	AACAACAAACAAACGACGACGACGACGCTCAACGATCAACGACAAACGACAAAC	168				
QY	1240	GGCAGCGAC	1248				
DB	167	AACGGCAAC	159				

RESULT 9  
CL973991  
LOCUS  
DEFINITION  
O81FC025416 Oryza sativa Express Library Oryza sativa (indica  
cultivar-group) genomic, genomic survey sequence.  
CL973991  
CL973991.1 GI:52402507  
GSS.  
Oryza sativa (indica cultivar-group)  
Oryza sativa (indica cultivar-group)  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 3069)  
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,  
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,

Wong, G.K.S., Deng, X.W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and  
its comparison to Arabidopsis  
Unpublished (2004)  
Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.  
Location/Qualifiers  
1..3069  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:39946"  
/clone\_lib="Oryza sativa Express Library"  
/note="Oryza sativa exon trapped genomic sequences"

## ORIGIN

		Query Match	3.4%;	Score 84;	DB 9;	Length 3069;	
		Best Local Similarity	42.8%;	Pred. No. 6.7e-05;			
		Matches 536;	Conservative	0;	Mismatches 710;	Indels	6; Gaps 2;
QY	37	ACCAGCGCCAACTCTGATGAGCGCAGCAACTTCAAGGGCCCCAAGCGCATCATCAAG	96				
DB	625	AGCGGATCGGCAAGACGACGCTGGCGAGTCTCTACCGGAGGAACAGCATCCGGAAG	684				
QY	97	TGTTTCAACTGCGGCAAGGAGGCGCACATGCCCCGCAACTGCCCGCCCCCCCCCAAGAAG	156				
DB	685	CATTTCGCTGCACGCGGTGGTCCAGTCCCCCGAAGTTCGCTTCGCGGACCTCCTC	744				
QY	157	GGTGTGGAAGTGGCGGCAAGAGGCGCACAGATGAAGACTGACCGAGCCGAGGCC	216				
DB	745	CACAGCTCATCGCCAGGCGCAACGCTCTCTCGCGGACGACAGCTCGCGCGGAGGC	803				
QY	217	ACTTCTTCGCGAGGACCTGGGCTTCCCGGAGGCAAGSCCGCGAGTTCGCCAGCGAG	276				
DB	804	--CGCGACGACGCTGCGCGCGGGTGGATGGAGCGGAGAGGCTCTGTCAAGCG	861				
QY	277	CAGAACCAGGCGCAACAGCCCCACACAGCGGAGCTGAGTGGCGGCGGCGCAACACCCCGC	336				
DB	862	CGCTTCGCGCGCGCTCCAGGCGAAGAGTACTCTGTGTGCTCGACGACGCTCGGAGC	921				
QY	337	AGCGAGGCGCGCGCGAGCGCCAGGCGACCTTGAACTTCCCCCAGATCACCTGTGGCAG	396				
DB	922	AGGAGGAGTGGGCGCTTCTTCTCGCGCGCTCCCCGCGGCGCTCGCGCGAGCTGCGTG	981				
QY	397	CGCCCCCTGTGAGCATCAAGTGGCGGCGCAGATCAAGGAGGCTCTGTGGACACCGGC	456				
DB	982	CTGTCAAGCAGCAGGTCAAGATCAGCGAGGATCGAGGACACCGTGGCGCGCGCGC	1041				
QY	457	GCGCAGCACACCGTGTGGAGGAGATGAGCTTCCCGCGCAAGTGGAAAGCCCAAGATGATC	516				
DB	1042	GGCGTGTGCGCACACGAGGAGCTCGGGAAGCTCGCGCGGAGCAGGCGCGCGAGCTGTT	1101				
QY	517	GGCGGATCGCGCGCTTCAAGTGGCGCAGTACGACAGATCTGTATCGAGATCTGC	576				
DB	1102	CGCGGAGGCTGTACGCGCCACGCGGAGCGGAGCCGGAAGAGATGGAGAGCTCAAGAGC	1161				
QY	577	GGCAAGAAGGCCATCGGCAACCGTGTGTCGGCCCCCACCCCTGAACATCATCGGCGCGC	636				
DB	1162	CTGTGATCTCATGAGGAGGCTCAACCTGCGCTCAACATGCTATGCTCGCGGA	1221				
QY	637	AACATGCTGACCGAGCTGGCTGCACCTTGAACTTCCCCATAGCCCCCATCGAGACCGTG	696				
DB	1222	CTCTTGAGGTCCAAAGAAGGAGGATGAATGGAGGCGCTGATCCATAGCTCGATGATACG	1281				
QY	697	CCGTGAAGTGAAGCCCGGCGATGGAGGCGCCCAAGGTGAAGCAGTGGCCCCCTGACCGAG	756				
DB	1282	CCACCCCGACATCGCGCGGAGCAGACCGCGCGGAGCGGAGCCGAAGAGGACAGCCTC	1341				

Qy 757 GAGAAGATCAAGGCGCTTACCGCCATCTGCAGGAGATGGAGAAGAGGGCAAGATCAC 816  
|||  
Db 1342 GATGATAGCAGAGCGGCGCCGACATCGACGAGCAAAACGAAGAAGAGCAAGAGGAAG 1401  
|||  
Qy 817 AAGATCGGCGCCGAGAACCCCTACACACACCCCGTGTTCGCATCAAGAGAGGACAGC 876  
|||  
Db 1402 CAGATGGGACATCGATGGAACAAGATCTGACGGTGTGTCATGGAGAGACTTCCGACGAT 1461  
|||  
Qy 877 ACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936  
|||  
Db 1462 CTGAAGCGGTCTTCTCTACTTCGCGGGTTACCGCGAGACGCGGATCTGCGCGGG 1521  
|||  
Qy 937 GAGTGCAGCTGGGATCCCCACCCCGCGGCTGAAGAAGAAGAGAGGCTGACCGTG 996  
|||  
Db 1522 AAGCTGTCTCGGCTGTGGGTGCGCGAGGGTTCTATGAGGCGGAAGAACGGGACAGCGTG 1581  
|||  
Qy 997 CTGACGTGGCGGACGCTACTTTCAGGCTGCGCTGGACGAGGACTTCCGCAAGTACAC 1056  
|||  
Db 1582 GAGGAGCAGCGCGAGGAGTGCTCATAGGAGCTCATCTCCCGGTCTGCTCCAGCTCGTC 1641  
|||  
Qy 1057 GCCTTACCATCCCGCAGCATCAACAAGAGACCCCGGCTATCCGATCAAGATCAAGCTG 1116  
|||  
Db 1642 GAGAGGACGCGCGGCGGAGGTCGCGCGCTCAGCATCCACCAAGCGGTCTCGAC 1701  
|||  
Qy 1117 CTGCCCCAGGCTGGAGGGGAGCGCCGAGCATCTTCCA---GAGCAGATGACCAAGATC 1173  
|||  
Db 1702 TTCTCCAGGCGGAGGCGCGACACCAACTTCTTCCACGTCACAGCGCGCGCGCGGC 1761  
|||  
Qy 1174 CTGGAGCCCTTCCGCGCCCGAACCCGAGATCTGATCTACCATGATGAGCAGCTG 1233  
|||  
Db 1762 CTCTCAACGCGCGCGCGCGCTCTGCGCTCGCAACACTACGACCGACCTCGCC 1821  
|||  
Qy 1234 TAGCTGGGACGAGCTGGAGATCGGCGAGCAGCACCGCGCGCAAGATCGAGGAGC 1285  
|||  
Db 1822 GTCATGCTCGAGGCGCCAGCTCACACCTCTCTCTCGGACATCCCGAGC 1873  
|||

## RESULT 10

CL971508  
LOCUS CL971508  
DEFINITION OaIFC021485 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.  
CL971508  
CL971508.1 GI:52397596  
GSS.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Oryza sativa (indica cultivar-group)  
Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

## REFERENCE

## AUTHORS

Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G.K.S., Deng, X.W. and Wang, J.  
An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis  
Unpublished (2004)

## JOURNAL

## COMMENT

Contact: Chen Chen  
Department of Bioinformatic  
Beijing Institute of Genomics  
Chinese Academy of Sciences, Beijing 101300, China  
Tel: 86-10-80481559  
Fax: 86-10-80488676  
Email: chenchen@genomics.org.cn  
Rice genomic sequence.  
Class: exon-trapped.

## FEATURES

## source

1..1941  
/organism="Oryza sativa (indica cultivar-group)"

/mol\_type="genomic DNA"

/db\_xref="taxon:39946"

/clone\_lib="Oryza sativa Express Library"

/note="Oryza sativa exon trapped genomic sequences "

## ORIGIN

Query Match 3.4%; Score 83.8; DB 9; Length 1941;  
Best Local Similarity 42.9%; Pred. No. 7e-05;  
Matches 769; Conservative 0; Mismatches 992; Indels 30; Gaps 6;  
Qy 267 CCCAGCAGCAGAACCGCGCAACAGCCCAACAGCCGCGAGCTGCAAGTGGCGGCGA 326  
Db 165 CGCGCCCAAGAACCGAGTCCGATGAACCCCAACACACCGCTTTCGATGGAAGCGGTT 224  
Qy 327 CAACCCCGCAGCAGCGCGCGCGCGAGCGCCAGCGCACCTGAACTTCCCCAGATCAC 386  
Db 225 GATTGGTAGAGGTTCTCCGACCCGTCGTCAGAGTGACATGAAGTATGGCGGTTCAA 284  
Qy 387 CCTGTGCAGCGCCCTCTGTGTAGCATCAAGGTGGCGGCGCAGATCAAGGAGGCGCTCT 446  
Db 285 GGTGTCTCTGCGCCCGCGCAGCCGATGATCGTCCAGTACAAGGGGAGAGAA 344  
Qy 447 GGAACCGCGCGCAGCAGCACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAAAGCC 506  
Db 345 GCAGTTCGCGCGGAGGAGATCTCGTCATGCTGATCAAGATGAGGGAGATCGCGGA 404  
Qy 507 CAAGATGATCGGCGCATCGCGGCTTTCATCAAGTGGCGGCGCAGTACGACAGATCTGAT 566  
Db 405 GGCTTACTCTGGTCTGTCATCAAGAACGCGGTGTCAACGTCCTCCGCGCTACTTCAACGA 464  
Qy 567 CGAGATCTGCGGCAAGAGCGCATCGGCACCGTCTGATCGGCGCCCGCCCGCTGMAACAT 626  
Db 465 CTGCGAGAGCAGGCCACCAAGAGCGCGGCTCATCGCGGGTCAACGTGATGCGGAT 524  
Qy 627 CATGCGCGCAACATGCTGACCCAGCTGGGTGTCACCTTGAATTTCCCGCATCAGCCCAT 686  
Db 525 CATCAACGAGCCACCGCGCATCGGCTCGACAAAGAGGCGAGCAGCAG 584  
Qy 687 CGAGACGTCGCGTGAAGCTGAAGCCCGGATGGACGGCCCGCCCAAGGTGAAGCATGGCC 746  
Db 585 CGCGAGAAGAACGTCTCATCTTTGACCTCGCGCGCGCACGTTTCGATGTGTGCTGCT 644  
Qy 747 CTTGACCGAGAGAAAGTCAAGCGCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGG 806  
Db 645 CACCATCGAGAGGGGATCTTCGAGGTCAAGGCCACCGCGCGGACACCCACCTCGCGGG 704  
Qy 807 CAAGATCACCAAGATCGGCGCCCGAGAACCCCTAC---AACACCCCGTGTTCGCCATCAA 863  
Db 705 CGAGGACTTCGACAAACCGCATGTGTGAACACTTCGTGACAGAGTTCAAGCGCAAGAACAA 764  
Qy 864 GAAGAGGACAGCACCAAGTGGCGGCAAGCTGTGGACTTTCGCGAGCTGAACAAGCGCAC 923  
Db 765 GAAGGACATCAGCGGCAACCCCGCGCTCGCGCGCTCAGGACGCGGTGCGAGCGCGC 824  
Qy 924 CCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCCACCCCGCGCTGAAGAAGAGAGAA 983  
Db 825 CAAGCGCACCTGTCTGTCGAGCGGCGAGACACCATTTAGATCGACTCTGCTGTACGAGGG 884  
Qy 984 GAGCGTGACCGTGTGGACGTGGCGCGAGCGCTACTTTCAGCGTGGCCCTGGACGAGGACTT 1043  
Db 885 CATCGACTTCTACACAGCATCACAGGCGCGGTTTCGAGGAGCTCAACATGGACCTGTT 944  
Qy 1044 CCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAGAGACCCCGCGCATCCGCTA 1103  
Db 945 CCGCAAGTGTGATGAGCGCCCGTGGAGAGTGCCTCCGCGAGCCCAAGATGGACAAGAGCAG 1004  
Qy 1104 CCAGTACAAGGTGTGCCCGCAGGGCTGAAGAGGCGAGCCCGAGCATTTTCCAGAGCAGCAT 1163  
Db 1005 GTGCGACGCTGTGCTC-----GTGCGGGGTCTCACCGCATTCGCCAGGTGCA 1055  
Qy 1164 GACCAAGATCTGGAGCCCTTTCGCGCGCCGCAACCCCGAGATCGTGTATCTACCAAGTACAT 1223  
Db 1056 GCAGCTGCTCCAGGACTTCTTCAAGCGCAAGGAGCTCTGCAAGAGCATCAACCCCGACGA 1115  
Qy 1224 GGAAGACCTGTACGTGGGCGAGCACTTGGAGATGGCGCAGCACCGCGCCCAAGATCGAGGA 1283  
|||

```
Db 1116 GGCGGTGGCTACGGCGCGCGCTGCAGGCGCGCATCTTGAGCGCGAGGGAACAGGAA 1175
QY 1284 GCTGGCAAGCACCTGCTGGCTGGGCTTACACACCCCGCAAGAAGCACCAGAGGA 1343
Db 1176 GGTGC---AGGACCTGCTGCTGCTGACGCTCACCGGCTGTCCCTCGGCTCGAGACCGC 1232
QY 1344 GCCCCCTTCTCTGATGGCTACGAGCTGCACCCCGCAAGTGACC---GTGCGAGCC 1400
Db 1233 CGCGGGGCTCATGACCGTCTATCCGAGGAACACACCATCCCGACCAAGAGAGCA 1292
QY 1401 CATCGAGCTGCCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGTGGGCAA 1460
Db 1293 GGTCTTCTCCACCTACTCCGACAAACAGCCCGCGCTCTGATCCAGTGTACGAGGGCGA 1352
QY 1461 GCTGAATCTGGCGAGCAGATCTACCCCGGCATCAAGTGGCGCAGCTGTGCA----- 1513
Db 1353 GCGCGCAGGACCAAGGACAAACCTCTCTCGGCAAGTTTCGAGCTCTCCGGGCATCCACCC 1412
QY 1514 --AGCTGCTGCGGCGCGCAAGGCCCTGACCGACATCTGTGCCCTGACCGAGGAGCGGA 1571
Db 1413 GCGCGCGCGGCGTGCCTCCAGATCAACGTCTGCTTGCATCGACGCCAAGCGGATCCT 1472
QY 1572 GCTGGAGCTGGCGGAGAACCGGAGATCTTGCAGGAGCCCGTGCACCGGCTGTACTACGA 1631
Db 1473 GAACGTGTCCGCGAGGACAAAGACGACGCGGCGAGAAACAAGATCAACATCAACCAAGCA 1532
QY 1632 CCCAGCAAGGACCTGTGTGGCGAGATCCAGAGAGGAGGCGACGACGAGTGGAACCTACCA 1691
Db 1533 CAAGGGGCGGCTGAGCAAGAGGAGATCGAGAAGATGGTSCAGGAGGCGGAGAGTACAA 1592
QY 1692 GATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGC 1751
Db 1593 GCGGAGGAGCAGAGACAAAGAAAGAGTGAGAGCGCCAGAAAGCCCTCGAGAACTACGC 1652
QY 1752 CCAACCAACAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGGAGAGCAT 1811
Db 1653 CTAAACATCGCAACACCATCAAGAGCAGAGAAGATCGCCTCCAAGTCTCGCCGCGAGCA 1712
QY 1812 CGTGATCTGGGGAAGACCCCAAGTTCCGCTGCCCTCCAGATCCAGAGGAGACTGGGAGAC 1871
Db 1713 CAAGAAGAGGATCGAGGAGCCCATCGATGCGCGCCATCAGCTGGCTCGACACCAACAGCT 1772
QY 1872 CTGGTGACCGCACTACTGGCAGCGCCACTGTGATCCCGAGTGGGAGTTCTGTGAACACCCC 1931
Db 1773 CGCGGAGCGCAGATTTCAGAGACA---AGATGAAGGAGCTGGAGGGCATCTGCAACCC 1829
QY 1932 CCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAGAGGACCCATCATCGGCGCCGAGACCTT 1991
Db 1830 GATCATCGCAAGATGTACAGGGCCCGCGCGCATATGGCGGAGGATGGACGAGGA 1889
QY 1992 CTAGTGTGACGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGGCTA 2042
Db 1890 CGCGCGCGGCGGCGAGCGCGCAGCGCCCAAGATCGAGGAGTTCGACTA 1940
```

## RESULT 11

```
BM320864 1132 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
sequence.
```

BM320864 1 GI:18055270

## ACCESSION

BM320864.1

## VERSION

BM320864.1

## KEYWORDS

Mastigamoeba balamuthi

## SOURCE

Mastigamoeba balamuthi

## ORGANISM

Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.

## REFERENCE

1 (bases 1 to 1132)

Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W.,

Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and

Philippe, H.

The analysis of 100 genes supports the grouping of three highly

divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba

JOURNAL  
MEDLINE  
PUBMED  
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)  
21819461  
11830664  
Contact: Muller Miklos  
Laboratory of Biochemical Parasitology  
The Rockefeller University  
1230 York Avenue, New York, NY 10021, USA  
Email: mmuller@rockvax.rockefeller.edu  
Insert Length: 1132 Std Error: 0.00  
POLYA=No.

FEATURES  
source

1. .1132  
/organism="Mastigamoeba balamuthi"  
/mol\_type="mRNA"  
/strain="ATCC 30984"  
/db\_xref="taxon:108607"  
/clone\_lib="Mastigamoeba balamuthi lambda ZAP II Library"  
/note="syn: Phreatamoeba balamuthi"

## ORIGIN

```
Query Match 3.2%; Score 80.2; DB 4; Length 1132;  
Best Local Similarity 45.3%; Pred. No. 0.0003;  
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;  
QY 139 CGCGCCCCCGCAAGAGGCTCTGGAAGTGCAGGAGGAGGCGCCACCAAGATGAAGGAC 198  
Db 79 CGCGCGAGGGGCAAGACGACTTACCGGCGCGCCANCTGGTGATCCAGGACAAACAAG 138  
QY 199 TGCAACGAGCGGCAGGCCAACTTTCTTCGCGAGGACCTGGCCCTTCCCCCAGGGGCAAGGCC 258  
Db 139 TACAACAGGCCCCAAGTACCGCTTCGTC-----GTCCGCTTCAACAACAGGAGCATCGTC 192  
QY 259 CGCGAGTTCCTCCAGCGAGGAGACCGCGCAACAGCCCCACAGCGCGGAGCTGCAGGTG 318  
Db 193 TGCAGATCGCTTACCGCAAGATCGAGCGGACCAATCTCTCGCGCGCTACTGCGAC 252  
QY 319 CGCGGCGCAACACCCCGCAGCGAGGCGCGCGCGAGCGCGCAGGGCACCTTGAATTTCCCC 378  
Db 253 GAGCTCACCGCTTCGGCGTCAAGCTCGGCTACCAACTACGCCGCGCTACCGGACT 312  
QY 379 CAGATCACCTGTGGCAGCGCCCTTGTGTGAGATCAAGGTGGGCGGCGAGATCAAGGAG 438  
Db 313 GGCTCTGTCTGGCGCGCGTGTGCTGAAGAAGCTCAACCTCGACTTCAAGTACGAGGGT 372  
QY 439 GCCTCTGTGACACCGCGCGCGACACCGCTCTGAGGAGATGAGCTTCCCGGCAAG 498  
Db 373 GTCAAGAAGTCAACGGCGAGGACTAAGCTGAGAGCTCGACGAGGCGCGCGCGCTTC 432  
QY 499 TGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTATCAAGGTGCGCCAGTACGACAG 558  
Db 433 TTCAAGGCGCTGTCTCGACGCTCGGCTGTTCGCGACCTCGACTGGCGCGCGCTTTCGCC 492  
QY 559 ATCTGTATGAGATCTTGGCGCAAGAGGCGCATCGGACCGGTGTGTGATCGGCGCGCGCCC 618  
Db 493 GCCTCAAGGGCATGTGCGACGCGCGCGTCAACGTCCCC---CACAGCGAGACCGCGCTTC 549  
QY 619 GTGAACATCATTCGCGCGCAACATGTCACCCAGCTGGGCTGCACCTTGAATTTCCCCATC 678  
Db 550 GTCGGGTTCAACGGCGACAAAGAGGAGCTCAAGCGCGGCTCTCTCGCAAGTATCATCTTC 609  
QY 679 AGCCCATCGAGACCGTGGCCCGGTGAAGCTGAAGCCCGGCGATGGACGGCGCCCAAGAGTGAAG 738  
Db 610 GGCGGCCACGTCCCGGTATATGAAGCTCTCAAGGAGCAGGACGCGCGCGCTTCGAC 669  
QY 739 CAGTGGCCCTGTACCGAGGAGAAAGATCAAGGCCCTGACCGCCATCTGCGGAGGAGATGAG 798  
Db 670 CGCCAGTTCTCGCGCTACGCCAAGAGGAGGTGTCCCGCGGACATGCTTCGAGAGAGATCTAC 729  
QY 799 AAGGAGGCGAAGATCAACCAAGATCGGCCCGCGAGAACCCCTACAACACCCCGCTTTCGCC 858  
Db 730 ACCGAGGCCCAAGCAGATCCCGCGCGGACCCCGACCTTCGTCCCCAAGCGCGCTCGAAG 789  
QY 859 ATCAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTTCGCGGAGCTTGAAACAAG 918
```

```
Db 790 CCCAGGCGCCGACCCAGCCTGGGGCAAGCGCAGCTGACGTACACGAGCGCAAG 849
Qy 919 CGCACCCAGGACTTCTGGAGGTGTCAGCTGGGCTATCCCGCCCGCC 966
Db 850 AACCGGCTGCGCCAGAGAAGTCCGCTGGCTACCCCGACGCCCC 897

RESULT 12
BM320900 1165 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller.0.353 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
sequence.
ACCESSION BM320900
VERSION BM320900.1 GI:18055306
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
REFERENCE 1 (bases 1 to 1165)
AUTHORS Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1165 Std Error: 0.00
POLYA-No.
Location/Qualifiers
source 1. 1165
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"

ORIGIN
Query Match 3.2%; Score 79.6; DB 4; Length 1165;
Best Local Similarity 45.1%; Pred. No. 0.00039;
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;

Qy 90 CATCAAGTGTTCAACTCGCGCAAGAGGGCCACATCGCCCGCAACTCGCGCGCCCGCC 149
Db 46 CGTCAAGNACAAGCGGTACTTCAAGCGCTTCCAGACCCAGTTCCGTGCGCCCGGAGG 105
Qy 150 CAAGAAGGGTGTGGAAGTGGCGCAAGAGGGCCACAGATGAAGGACTGCACCGAGCG 209
Db 106 CAAGACGGACTACCGCGCGGCCANCTGGTGATCCAGGACAAGAAACAAGTACACAGCC 165
Qy 210 CCAGGCCAACTTCTTCGCGAGGACCTGGCTTCCCGGCAAGGCCCGCGAGTTCCC 269
Db 166 CAAGTACCGTTTCTGTC-----GTCCGCTTCCACCAAGGGGACATCTCTGCCAGATCGC 219
Qy 270 CAGCGACGAGAACCGCGCCCAACAGCCCGCCAGCGCGAGCTGCGAGTGGCGCGGCAAA 329
Db 220 CTAGCCCAAGATCGACGGCGNACATCTCTCGCGCGCGCTACTCTCGCAGAGCTCACCG 279
Qy 330 CCCCCGAGCGAGCGCGCGCGCGCGCCAGCGCCACCTGAACTTCCCGCAGATCACCT 389
Db 280 CTTGCGGCTCAAGTCTCGGCTGACCAACTACCGCGCGCGCTACCGGACTGGCTGTGCT 339
Qy 390 GTGGCAGCGCCCTGTGTGAGCATCAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 449
Db 340 GGCCCCGCGTGTGCTGAAGAAGCTCAACCTCTACTCCAAAGTACGAGGTGTCAAGAAGGT 399
```

```
Qy 450 CACCGGCGCCGACGACACCGTGTCTGGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAA 509
Db 400 CAACGGCGAGGACTTACAACGTCGAGGAGCTCGACGACGGGCGCCCGTTCGAAGGCC-- 457
Qy 510 GATGATCGGCGCATCGGCGCTTCATCAAGGTGGCCAGTACGACAGATCTCTGATCGA 569
Db 458 --TGCTCGACGTGCGCTTGGTCCGACCTCGACTGGGCGCCCGCGTGTTCGCGCGCCCTCAA 515
Qy 570 GATCTGGCGGCAAGAGCCATCGGCACCGTCTGATCGGCGCCCGCCCGCGTGAACATCAT 629
Db 516 GGGCATGTGCGACGGCGGCTCAAGTCCCGCCACAGCGAGACCGCTTCTGTCGGTTCAA 575
Qy 630 CGGCGGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCATCATAGCCCCATCGA 689
Db 576 CGGCGCAAGAAGGAGCTCAACGCGCGGCTCTCCGCAAGTACATCTTTCGGCGGCGCAGT 635
Qy 690 GACCGTCCCGTGAAGCTGAAGCCCGCATCGAGCGGCCCGCCCAAGGTGAAGCAGTGGCCCT 749
Db 636 CGCGCGGTACATGAAGCTCTTCAAGGAGCAGGACGCGCGCGCTTCGACCCGCCAGTTCTC 695
Qy 750 GACCGAGGAGAGATCAAGGCCCTTGACCGCCATCTGCGAGGAGATGGAAGAAGGAGGCAA 809
Db 696 GCGTACGCCAAGAGGGTGTACCGCGGACATCTCGAGAAGATCTACACGAGGCCCA 755
Qy 810 GATCACCAGATCGGCGCGGAGAAACCCCTCAACACCCCGCTGTTCGCCATCAAGAAGAA 869
Db 756 -----CAAGCAGATCGCGCGCGACCTTCTCCCAAGCGCGCTCGAAGCCCGA 809
Qy 870 GGACAGCACCAGTGGCGCAAGCTGGTGGACTTCGCGGAGCTGAACAGCCGCCACCGA 929
Db 810 GGGCGCCAAAGCCCAAGCCTGGGGCAAGCGCAGCTGACGTACCACGAGCGCAAGAACC 869
Qy 930 CTTCTGGAGGTGCGAGCTGGGCATCCCGCCCGCGCTGAAGAGGAAGAAGAGCGT 989
Db 870 CGTGGCCCAAGAGGTTCGCTGGGTACCCCGAGCGCCCGCCAGAGCCAGTAAATTCC 929
Qy 990 GACCGTGTGAGCGTGGCGGACGCTTCTTTCAGCGTGCCTTCGACGAGG 1039
Db 930 CGGTGCCCTGTACACCGCCATCTCCCGCTGGCGGTTCGCTGTGCGG 979
```

```
RESULT 13
BM321430 867 bp mRNA linear EST 03-JAN-2002
LOCUS rockefeller.0.1153 Mastigamoeba balamuthi lambda ZAP II Library
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein S4, mRNA
sequence.
ACCESSION BM321430
VERSION BM321430.1 GI:18055836
KEYWORDS EST.
SOURCE Mastigamoeba balamuthi
ORGANISM Mastigamoeba balamuthi
REFERENCE 1 (bases 1 to 867)
AUTHORS Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
Philippe,H.
TITLE The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
MEDLINE 21819461
PUBMED 11830664
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 867 Std Error: 0.00
POLYA-No.
Location/Qualifiers
source 1. 867
/organism="Mastigamoeba balamuthi"
```





```
QY 1116 GCTGCCCCAGGGCTGAAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175
Db      |||||
Db 1002 CAAGGCCAGATCCAGATGCTGCTCTCGTGGAGGCTCCACCGTATCCCCAAGGTGCA 1061
QY 1176 GGAAGCCCTTCCGGCCCGCAACCCCGAGATCGTGATCTACAGTACATGAGACGACCTGTGA 1235
Db      |||||
Db 1062 GCAGCTCTGTCAGGACTTCTTCAACGGGAAGGAGCTCTGCAAGAGCATCAACCCCGACGA 1121
QY 1236 CPTGGGCAGCGACTGTGAGATCGGCAGACCGCGCCAGATCGAGGAGCTGCGCAAGCA 1295
Db      |||||
Db 1122 GGCCTGTGGCGTACGGGCGCGCTCCAGGCGCGCATCTCAGCGGGAGGGCAACCAAGAG 1181
QY 1296 CTTGCTGCGCTGGGGCTTACCAACCCCGCACAAAGAGCACAGAGAGCCCCCTTCTCT 1355
Db      |||||
Db 1182 GGTGCAGGACTGCTCTCTCTGACGTCACCGCTCTCGCTCGGTTTGAGACGGCCGG 1241
QY 1356 GTCGATGGGTACAGAGTGCACCCCGACAAGTGGAGCGGTGCAGCCCATCGAGCTGCCCGA 1415
Db      |||||
Db 1242 GGGTGTCAATGACCGTGTGATCCGAGGAACACCAACCTCCCAACCAAGAAGGAGCAGGT 1301
QY 1416 GAAGGAGAGTGAACCGTGAACACATCCAGAGAGTGGTGGGCAAGCTGAACTGGGCCAG 1475
Db      |||||
Db 1302 CTTCTCCACCTACTCCGACAAACAGCCCGCGCTGCTCATCCAGGTGTACGAGGGCGAGAG 1361
QY 1476 CCAGATCTACCCCGGATCAAGGTGGCCAGCTGTGCAAGCTGCTGGCGGCGCCAAGGC 1535
Db      |||||
Db 1362 GACGAGGACCAAGGACAACTGCTCGCAAGTTGAGCTCACCGGCATCCGCGCGC 1421
QY 1536 CTTGACCGCATCGTGCCTGTGACCGAGGAGCGGAGCTGGAGCTGGCCCGAGAACCCGGA 1595
Db      |||||
Db 1422 GCCGAGGGGGTGCCTCCAGATCAACGTGACGTTTCGACATCGACGCGAAACGGCATCTGAA 1481
QY 1596 GATCCTCGCGAGCCGTGCACGGCGTGTACTAGCA-----CCCCAGCAAGGACCTGGT 1649
Db      |||||
Db 1482 CGTGTGCGCGGAGGACAAGACGACGGGGAAGAAGAACAGATCACCATCAACCAACGACAA 1541
QY 1650 GGCCGAGATCCAGAAGCAGGGCCACACACAGTGGACCTTACCAGATCTTACCAGGAGCCCTT 1709
Db      |||||
Db 1542 GGGCGGCTGAGCAAGGAGAGATCGAGCGGNTGGTGCAGGAGGCCGAGAGTACAAGGC 1601
QY 1710 CAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACACCAACGACGTGAA 1769
Db      |||||
Db 1602 CGAGGACGAGCAGTGTGCGGCAAAAGGTGGAGGCGCGCACGCGCTGGAGAACTACGCGTA 1661
QY 1770 GCAGCTGACCGAGCCGTGCAGAGATCGCCATGGAGAGCATCGTGATCTGGGGCAGAC 1829
Db      |||||
Db 1662 CAACATGCGCAACACGCTGCGGCAAGAGATGCGCTCCAAGCTCCCGCGCAGCAGCAA 1721
QY 1830 CCCCAAGTTCGGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTG 1889
Db      |||||
Db 1722 GAAGAGATCGAGGACGCCATCGAGGATGCCATCAAGTGGCTCGACGGCAACCAAGCTCGC 1781
QY 1890 GCAGGCCACTGTGATCCCGAGTGGGAGTTCTGTGAACACCCCCCTCTGTGAAGCTGTG 1949
Db      |||||
Db 1782 CGAGGCGGACGAGTTCGAGGACAAGATGAAGGAGCTCGAGAGCTCTGCAACCCGATCAT 1841
QY 1950 GTACCACTGGAGAGGAGCCCATCATCGCGCGGAGACCTTCTACGTGACCGGCGCGC 2009
Db      |||||
Db 1842 CTCAAAGATGTACCGGGCGCGCGCTGCTGCGGGGATGGAGCAAGACCGCCCCAA 1901
QY 2010 CAACCGGAGACCAAGATCGGCAAGGCCGCGC 2040
Db      |||||
Db 1902 CGGCGTGCCTGGCACCGGCGGTGGCAGCGGC 1932
```

Search completed: June 2, 2005, 02:53:19  
Job time : 7886.14 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 03:13:23 ; Search time 34 Seconds  
(without alignments)  
10815.332 Million cell updates/sec

Title: US-09-610-313B-31  
Perfect score: 4573  
Sequence: 1 gtcagcaccatgcccga.....999ctagcaccggtgaattc 2463

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09610313/runat\_31052005\_155137\_15173/app\_query.fasta\_1.7893  
-DB=Issued Patents AA -QPMF=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CNG 1.101 @runat\_31052005\_155137\_15173 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3497	76.5	1005	2	US-07-743-357-1
2	3493.5	76.4	1003	2	US-07-743-357-9
3	3475	76.0	1015	3	US-08-463-210-9
4	3475	76.0	1015	3	US-09-124-900-3
5	3475	76.0	1015	4	US-08-463-028-9
6	3475	76.0	1016	2	US-07-743-357-2
7	3472	75.9	1003	4	US-09-309-572-17
8	3472	75.9	1003	4	US-07-718-096-17
9	3469	75.9	1016	2	US-07-743-357-5
10	3468	75.8	1016	2	US-07-743-357-3
11	3465	75.8	1016	2	US-07-743-357-4
12	3459.5	75.7	1003	2	US-07-743-357-10

13	3456.5	75.6	1003	2	US-07-743-357-8	Sequence 8, Appli
14	3449	75.4	1004	2	US-07-743-357-7	Sequence 7, Appli
15	3183.5	69.6	1014	4	US-09-319-588C-6	Sequence 6, Appli
16	3141	68.7	913	2	US-07-743-357-22	Sequence 22, Appli
17	3119.5	68.2	1350	4	US-09-952-060-35	Sequence 35, Appli
18	3078	67.3	913	2	US-07-743-357-6	Sequence 6, Appli
19	2822.5	61.7	875	4	US-09-952-060-6	Sequence 6, Appli
20	2817	61.6	850	4	US-09-952-060-2	Sequence 2, Appli
21	2798.5	61.2	875	4	US-09-952-060-8	Sequence 8, Appli
22	2793	61.1	850	4	US-09-952-060-4	Sequence 4, Appli
23	2789	61.0	562	3	US-09-117-217-14	Sequence 14, Appli
24	2789	61.0	562	3	US-09-735-487-14	Sequence 14, Appli
25	2758	60.3	560	4	US-10-205-641-1	Sequence 1, Appli
26	2507.5	54.8	1018	4	US-09-206-551-46	Sequence 46, Appli
27	2483.5	54.3	1016	4	US-09-625-972-24	Sequence 24, Appli
28	2259	49.4	441	4	US-09-991-258-16	Sequence 16, Appli
29	2212	48.4	1055	2	US-08-659-251-5	Sequence 5, Appli
30	2212	48.4	1055	3	US-09-256-490-5	Sequence 5, Appli
31	2212	48.4	1055	5	PCT-US96-11445-5	Sequence 5, Appli
32	2193.5	48.0	3080	6	5223423-4	Patent No. 5223423
33	2193.5	48.0	3080	6	5223423-4	Patent No. 5223423
34	2154	47.1	427	3	US-09-690-265-1	Sequence 1, Appli
35	1572	34.4	327	3	US-08-679-493A-72	Sequence 72, Appli
36	1555	34.0	327	3	US-08-679-493A-71	Sequence 71, Appli
37	1552	33.9	314	1	US-08-589-446-6	Sequence 6, Appli
38	1552	33.9	314	1	US-08-444-882-6	Sequence 6, Appli
39	1552	33.9	314	2	US-08-389-459A-6	Sequence 6, Appli
40	1552	33.9	314	3	US-08-987-867A-6	Sequence 6, Appli
41	1311	28.7	1150	3	US-09-238-303-9	Sequence 9, Appli
42	1311	28.7	1150	4	US-09-946-239-9	Sequence 9, Appli
43	1309	28.6	275	6	5252477-3	Patent No. 5252477
44	1309	28.6	275	6	5252477-3	Patent No. 5252477
45	1297	28.4	261	6	5320958-3	Patent No. 5320958

#### ALIGNMENTS

#### RESULT 1

US-07-743-357-1

; Sequence 1, Application US/07743357

; Patent No. 5858646

; GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: KIRBY EADES GALE BAKER

; STREET: Box 3432, Station D

; CITY: Ottawa

; STATE: Ontario

; COUNTRY: Canada

; ZIP: K1M 1H8

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/743,357

; FILING DATE: 21-AUG-1991

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA90/00062

; FILING DATE: 23-FEB-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Gale, Edwin J.

; REGISTRATION NUMBER: 28,584

; REFERENCE/DOCKET NUMBER: 30924-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (613) 237-6900

; TELEFAX: (613) 237-0045

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1005 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHEICAL: NO  
 ; FRAGMENT TYPE: internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; STRAIN: HXB2  
 ; US-07-743-357-1

Alignment Scores:  
 Pred. No.: 1,22e-208 Length: 1005  
 Score: 3497.00 Matches: 655  
 Percent Similarity: 95.56% Conservative: 34  
 Best Local Similarity: 90.85% Mismatches: 26  
 Query Match: 76.47% Indels: 6  
 DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x US-07-743-357-1 (1-1005)

```

QY 220 TTCTTCGGCAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGCAGCGAGCAG 279
DB 2 PhePheargGluAspLeuAlaPheLeuGlnGlyAlaArgGluPheSerSerGluGln 21
QY 280 AACCGGCGCAACGCCCCACAGCGCGAGCTGAGGTG-----CGCGCGCAACACCCC 333
DB 22 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 41
QY 334 CGCAGCGAGGCGCGCGCAGCGCCAGGCGCACCTCG-----AACTTCCCGCAGATCACCC 387
DB 42 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 61
QY 388 CTGTGGCAGCGCCCTGTGTGATCAAGGTGGCGCGCCAGATCAAGGAGGCGCTGCTG 447
DB 62 LeuTrpGlnArgProLeuValThrLeuLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 81
QY 448 GACACGCGCGCAGACGACCGTGTGGAGAGATGAGCTGCGCGCGCAGAGTGAAGCC 507
DB 82 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 101
QY 508 AAGATGATCGCGCGCAGTCCGCGCTTCATCAAGGTGGCGCGCAGTACGACACGATCCTGATC 567
DB 102 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle 121
QY 568 GAGATCTGCGCGCAAGAGCGCATCGGCACCGTGTGATCGCGCGCCCGCCCGTGAACATC 627
DB 122 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 141
QY 628 ATCGCGCGCAACATGTGACCCAGCTGGTGTGACCTGCACTTCCCGCATCAGCGCCCATC 687
DB 142 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 161
QY 688 GAGACGTCGCGTGAAGCTGAAGCCCGCATGAGCGCGCGCAGTGAAGTGAAGTGGCCCC 747
DB 162 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 181
QY 748 CTGACCGGAGAGAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAAGAGAGAGGCG 807
DB 182 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 201
QY 808 AAGATCAACCAAGATCGCGCGCGAGAACCCCTACAACACCCCGTGTTCGCGCATCAAGAAG 867
DB 202 LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 221
QY 868 AAGCAGACCAAGTGGCGCAGCTGGTGGCTTCGCGAGCTGACACAGCGCCACCCAG 927
DB 222 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgGln 241
QY 928 GACTTCTGGGAGGTGACGCTGGGCATCCCCCACCCCGCGCTGGAAGAGAAAGAGAGC 987

```

```

DB 242 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 261
QY 988 GTGACCGTCTGCACGCTGGCGCAGCGCTACTTTCAGGTCGCCCGCTGACGAGGATCCCGC 1047
DB 262 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 281
QY 1048 AAGTACACGCGCTTACCATCCCGCAGCATCAACACGAGACCCCGCGCATCCCGCTACCAG 1107
DB 282 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 301
QY 1108 TACAACGTGCTGCCCGCAGGCTGGAAGCGCAGCCCGCAGCATCTTCAGAGCAGCATGACC 1167
DB 302 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 321
QY 1168 AAGATCTGGAGCCCTTCGCGCGCCCGCAACCCCGAGATCGTGATCTACCAG-----GCC 1221
DB 322 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyThrGlnTyMetAsp 341
QY 1222 CCCTGTGCTGGCGCAGCAGCTTGGAGATCGCGCAGCAGCCCGCGCAGATCCGAGGAGCTG 1281
DB 342 AspLeuTyThrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 361
QY 1282 CGCAACGACCTGTGCTGCGCTGGGCTTTCACACCCCGCAGCAAGAACACAGAGGAGCC 1341
DB 362 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 381
QY 1342 CCCTTCTGTGGTGGCTACGAGCTGCGCAGCTGCGCAGCAAGTGGACCTGAGCCCGCATCGAG 1401
DB 382 ProPheLeuTrpMetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleVal 401
QY 1402 CTGCGCGCAAGAGGAGAGTGGACCGCTGAACGACATCCAGAAGCTGGTGGCGAAGCTGAAC 1461
DB 402 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 421
QY 1462 TGGCGCAGCCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCG 1521
DB 422 TrpAlaSerGlnIleTyThrProGlyIleLysValArgGlnLeuLysLysLeuLeuArgGly 441
QY 1522 GCCAAGCGCTGACCGACATCGTGGCGCTGACCGCAGGAGCGCGAGCTGGAGCTGGCGCG 1581
DB 442 ThrLysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGlu 461
QY 1582 AACCGCGAGATCTGCGCGAGCGCGCTGTGCTGTACTACCGACCCCGCAGCAAGGAGCCTG 1641
DB 462 AsnArgGluLeuLysGluProValHisGlyValTyThrAspProSerLysAspLeu 481
QY 1642 GTGCGCGAGATCCAGAAGCAGGCGCAGCAGCTGAGTGTGACCATCTACAGAGGAGCC 1701
DB 482 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTyThrGlnIleTyThrGlnGluPro 501
QY 1702 TTCAGAACCTGAGACCGCGAGTACGCGCAGATCGCGCAGCGCGCAGCGCGCAGCGCTG 1761
DB 502 PheLysAsnLeuLysThrGlyLysTyAlaArgMetArgGlyAlaHisThrAsnAspVal 521
QY 1762 AAGCAGCTGACCGAGGCGCTGCAAGAGATCGCCATCGAGAGCATCTGTGATCTGGGCGAAG 1821
DB 522 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 541
QY 1822 ACCCCCAAGTTCGCTGCGCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1881
DB 542 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTy 561
QY 1882 TGGCAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTG 1941
DB 562 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 581
QY 1942 TGGTACAGCTGGAAGAGGAGCCCATCATCGCGCGCGCAGACCTTCTACGTGGAGCGCGCC 2001
DB 582 TrpTyThrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyThrValAspGlyAla 601
QY 2002 GCCAACCGCGAGACCAAGATCGCAAGCGCGCTGCTGACCGACCGCGCGCGCGAGAG 2061
DB 602 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyThrValThrAsnArgGlyArgGlnLys 621

```



Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTGTGAGCCCTCCGCGCCCGCCACACCGAGATGCTGATCTACACAG-----GCCCCC 1224  
 Db 321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
 QY 1225 CTGTACCTGGCGCAGCAGCTCGAGATCGCGCAGCAGCCGCCCAAGATCCGAGAGCTGCGC 1284  
 Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360  
 QY 1285 AAGCACTGTGCTGGCTGGCTTACACACCCCGCAGCAAGACCAAGAGAGAGCCGCC 1344  
 Db 361 GluHisLeuLeuLysTyrGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
 QY 1345 TTCTGTGAGTGTGCTGAGCTGACCGCCCGCAGCAGTGGACCGCTGAGCCCATCGAGCTG 1404  
 Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
 QY 1405 CCCGAGAAGGAGAGCTGGACCGCTGAACGACATCCAGAGCTGCTGGCAAGCTGAACCTGG 1464  
 Db 401 ProAspLysGluSerThrThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
 QY 1465 GCCAGCCAGATCTACCCCGCGCATCAAGTGGCCGAGCTGTGCAAGCTGTGCGCGCGCC 1524  
 Db 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLysLysLeuLeuArgGlyAla 440  
 QY 1525 AAGCCCTGACCCACATCTGCTGCGCCCTGACCGAGAGCCGAGCTGGAGCTGGCGGAGAC 1584  
 Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460  
 QY 1585 CGCGAGATCTCGCGAGCCGCTGCGAGCGCTGTACTACACCCCGCAGAGAGCTGGTGGT 1644  
 Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480  
 QY 1645 GCCGAGATCCAGAGAGCGCCACGACCACTGAGCTTACAGATCTACAGAGAGCCCTTC 1704  
 Db 481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500  
 QY 1705 AAGACCTGAGACCGCGAGTAGCGCAAGATGGCAGATGGCAGCCCGCCACACACGAGCTGAG 1764  
 Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
 QY 1765 CAGCTGACCGAGCGCGTGAGAGATCGCCATCGAGAGATCTGTGCTGGGCAAGAC 1824  
 Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnLysSerIleValIleTrpGlyLysThr 540  
 QY 1825 CCCAAGTTCGCTGCCATCCAGAGAGAGACTGGGAGACCTGGTGGACCGACTACTGG 1884  
 Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTyrTrp 560  
 QY 1885 CAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGCTGAGAGCTGG 1944  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
 QY 1945 TACCGCTGAGAGAGAGCCATCATCGCGCCGAGACCTTCTACGTGAGCGCGCGCC 2004  
 Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 QY 2005 AACCGGAGACCAAGATCGCAAGCGCGGTACTGACCGACCGCGGCGCGCAGAGATC 2064  
 Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 QY 2065 GTGAGCTGACCGAGACCAACCAACGAGAGAGCTGCGAGCCATCCAGCTGGCCCTG 2124  
 Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
 QY 2125 CAGGACAGCGCGCAGAGTGAAATCATCGTACCGACGAGCCAGTACGCTTGGGCATCATC 2184  
 Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 QY 2185 CAGGCCCGCCGACAGAGAGAGAGCTGGTGAACCCAGATCATCGAGAGCTGATC 2244  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

QY 2245 AAGAAGCAGAGGTGTACCTGAGCTGGTGGCCGCCCAAGAGGATCGCGGCAACGAG 2304  
 Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
 QY 2305 CAGATCGACAAAGCTGTGAGCAAGGAGCTCCGCAAGGTCTGTCTGACGCGCATCGAT 2364  
 Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720  
 RESULT 3  
 US-08-463-210-9  
 ; Sequence 9, Application US/08463210  
 ; Patent No. 6001977  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHANG, Nancy T.  
 ; APPLICANT: GALLO, Robert C.  
 ; APPLICANT: WONG-STAAAL, Flossie  
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morgan & Finnegan, L.L.P.  
 ; STREET: 345 Park Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10154-0053  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/463,210  
 ; FILING DATE: 05-JUN-1995  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/693,866  
 ; FILING DATE: 23-JAN-1985  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/659,339  
 ; FILING DATE: 10-OCT-1984  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Serunian, Leslie A.  
 ; REGISTRATION NUMBER: 35,353  
 ; REFERENCE/DOCKET NUMBER: 2026-4193US2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 758-4800  
 ; TELEFAX: (212) 751-6849  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1015 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: HTLV-III  
 ; FEATURE:  
 ; NAME/KEY: Protein  
 ; LOCATION: 1..1015  
 ; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
 ; US-08-463-210-9  
 Alignment Scores:  
 Pred. No.: 2,82e-207 Length: 1015  
 Score: 3475.00 Matches: 654  
 Percent Similarity: 93.72% Conservative: 33  
 Best Local Similarity: 89.22% Mismatches: 28  
 Query Match: 75.99% Indels: 18  
 DB: 3 Gaps: 4  
 US-09-610-313B-31 (1-2463) x US-08-463-210-9 (1-1015)

QY 220 TTCCTTCGCGAGGACCTGGCTTCCCTCCCGAGGGCAAGGCCCGCGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 268 -----CCGACGAGCAGAACCGCGCCACAGACCCCGACCGC 303  
Db 21 ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg 40  
QY 304 CGCGAGCTGCAGGTG-----CGGGCGACAAACCCCGACGAGCGCGCGCGAGCGC 357  
Db 41 ArgGluLeuGlnValTTPGlyArgAspAenAenSerProSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGGCACCCCTG-----AATCTCCCGCAGATCACCTGTGGCGAGCGCCCTGGTGAGC 411  
Db 61 GlnGlyThrValSerPheAenPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGGTGGCGCGCCAGATCAAGGAGCGCTGTGGACACCGCGCGCGAGCACCGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100  
QY 472 CTGAGGAGATGAGCTGCCCGCGCAAGTGGAGCCCAAGATGATCGGCGGCATCGGGGC 531  
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120  
QY 532 TTCTCAAGGTGCGCCAGTACGACAGATCCTGATCGAGATCTCGGCAAGAGGCCATC 591  
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGCACCGTGTGATCGGCGCCACCGCGTGAACATCATCGGCGCCCAACATGCTACCCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAenIleIleGlyArgAenLeuLeuThrGln 160  
QY 652 CTGGGCTGCACCTGAACTTCCCATCAGCCCATCAGACCGTGGCCCGTGAAGCTGAAG 711  
Db 161 IleGlyCysThrLeuAenPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGGCGATGAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCC 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCCGCTCTCGAGGAGATGGAGAGAGGCGGAGATCACCAAGATCGGCGCCGAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTTACAAACCCCGCTTCCCATCAAGAGAGAGAGACAGACCAAGTGGCGCAAG 891  
Db 221 AsnProTyrAenThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGC 951  
Db 241 LeuValAspPheArgGluLeuAenLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCACCCCGCGCGCTGAAGAAGAGAGAGCGTGACCGTGGAGCTGGCGGAC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCAGCTGCGCCCTGAGAGGACTTCGCGCAAGTACCGCTTCCACCATCCCC 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACAGAGACCCCGCGATCCGCTACAGTACACAGTGTGCTGCCCGGCGTGG 1131  
Db 301 SerIleAenAenGluThrProGlyIleArgTyrGlnTyrAenValLeuProGlnGlyTrp 320  
QY 1132 AAGGCGAGCCCGCATCTTCCAGAGCAGCATGACCAAGATCTCGGAGCCCTTCGCGGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCCGAGATCGGTATCTACAG-----GCCCCCTGTACGTGGGCGAGCGACTG 1245  
Db 341 GlnAenProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

QY 1246 GAGATCGGCGAGCACCGCGCCAAAGATCGAGAGCTGCGCAAGACACCTCTCGCTGGCTGGGC 1305  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1306 TTCAACACCCCGCAAGAAGACACCAAGAGAGCCCTTCTCTGTGTGATGGGTACGAG 1365  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuLeuTrpMetGlyTyrGlu 400  
QY 1366 CTGACACCCGCAAGTGGACCGCTGAGCTCCATCGAGCTGCCGAGAGAGAGAGTGGACC 1425  
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1426 GTGAACACATCCAGAGCTGGTGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGC 1485  
Db 421 ValAenAspIleGlnLysLeuValGlyLysLeuAenTrpAlaSerGlnIleTyrProGly 440  
QY 1486 ATCAAGGTGGCGAGCTGTGCAAGCTGCGCGCGCGCAAGGCCCTGACCGACATCGTG 1545  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1546 CCGCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAACCCGAGATCTCTCGCGAGGCC 1605  
Db 461 ProLeuThrGluGluAlaGluLeuGluAlaGluAenArgGluIleLeuLysGluPro 480  
QY 1606 GTGACCGCGTGTACTACGACCCCGACAGACCTGTGTGGCGAGATCCAGAGAGGCC 1665  
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1666 CACGACAGTGGACCTACAGATCTACGAGAGCCCTTCAAGAACCTGAAGCCGCGAAG 1725  
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAenLeuLysThrGlyLys 520  
QY 1726 TACGCCAAGATGCGCACCGCGCCACCAACAGACGTGAAGCAGCTGACCGAGCGCTGCAG 1785  
Db 521 TyrAlaArgMetArgGlyAlaHisThrAenAspValLysGlnLeuThrGluAlaValGln 540  
QY 1786 AAGATCCCGATGGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATC 1845  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1846 CAGAAGGAGACCTGGGAGACCTGGTGACCGACTACTGGCAGGCGCACCTGATCCCGAG 1905  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1906 TGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGTGACCTGACCTGGAGAGAGGCC 1965  
Db 581 TrpGluPheValAenThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro 600  
QY 1966 ATCATCGCGCGAGACCTTCTAGCTGAGCGCGCGCCCAACCCGCGAGACCAAGATCGGC 2025  
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAenArgGluThrLysLeuGly 620  
QY 2026 AAGSCGGCTACGTGACCGAGCCGCGCGCGCAGAGATCTGTAGCTTACCGACCGACCAAC 2085  
Db 621 LysAlaGlyTyrValThrAenLysGlyArgGlnLysValValProLeuThrAenThrThr 640  
QY 2086 AACCAAGAAGACCGAGCTGCAGGCGCATCCAGCTGCGCCCTGCGAGGACAGCGGAGGTG 2145  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAenSerGlyLeuGluVal 660  
QY 2146 AACATCGTGACGACAGCCAGTACGCCCTGGGCGCATCATCCAGGCCCGCCGACAGAGC 2205  
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2206 GAGAGCGAGCTGTGAACACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACTCG 2265  
Db 681 GluSerGluLeuValAenGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2266 AGCTGGGTGCGCCCAAGAGGCGCATCGCGCGCAACCGAGCAGATCGCAAGCTGTGAGC 2325  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2326 AAGGGCATCCGCAAGGTGTCTTCTCGGACCGGCATCGAT 2364

```
Db 721 AlaGlyIleArgYsIleLeuPheLeuAspGlyIleAsp 733
|||||
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
552 CTGGGCTGCACCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGAGAGCTGAAG 711
:::
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValIysLeuLys 180
712 CCGGCATGAGCGGCCCAAGAGTGAAGAGTGGCCCTCACCAGGAGAGAGATCAAGGCC 771
|||||
Db 181 ProGlyMetAspGlyProLysValIysGlnTrpProLeuThrGluGluIleLysAla 200
772 CTGACCGCCATCTGCAGGAGATGGAGAGGAGGCGCAAGATCACCAAGATCGGCCCGAG 831
|||
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
832 AACCCCTACACACACCCCTGTCGCATCAAGAAGACAGACACCAAGTCGCCCAAG 891
|||||
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240
892 CTGGTGCATCTCCGCGAGCTGAACAGCGCACCAGGACTTCTGGGAGGTGCAGCTGGGC 951
|||||
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
952 ATCCCCACCCCGCGCTGAAGAAAGAGAGCGTGACCGCTGTGGACGTGGCGGCAC 1011
|||||
Db 261 IleProHisProAlaGlyLeuLysLysSerValThrValLeuAspValGlyAsp 280
1012 GCCTACTTCAGCGTCCCGTGGACGAGGACTCCGCAAGTACACCGCTTACCATCCCC 1071
|||||
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
1072 AGCATCAACAAAGAGACCCCGCATCCGCTACAGTACAACTGCTGCCCAAGGCTGG 1131
|||||
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
1132 AAGGCGAGCCCGAGCATCTTCAGACGACGATGACCAAGATCTGGAGCCCTTCGCGGCC 1191
|||
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
1192 CGCAACCCCGAGATCGTGATCTACCAG-----GCCCGCTGTACCTGGCGACGACCTG 1245
|||
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeu 360
1246 GAGATCGGCGACGACCGCGCAAGATCGAGAGCTCGCAAGCACCTGCTGCCTCGCGGC 1305
|||
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380
1306 TTCACACCCCGCAGCAAGAGACCAAGAGAGCGCCCTTCTCTGTGGATGGCTAGAG 1365
|||
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400
1366 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGAGC 1425
|||
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
1426 GTGAACGACATCCAGAAAGCTGTGGCAAGCTGAACTGGGCGCACCGACAGATCTACCCCGC 1485
|||
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
1486 ATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGCGCTGACCGACATCGTG 1545
|||
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
1546 CCCTTACCGAGGAGCGCGAGCTGGAGCTGGCGAGAACCGGAGATCTCTGCGCGAGGCC 1605
|||
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysGluPro 480
1606 GTCCACCGCGTGTACTACGACCCCAAGAGGAGCTGTGCTGCCCGCGAGATCCAGAGAGCGGC 1665
|||
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
1666 CAGCACAGTGGACCTTACAGATCTACAGAGGAGCCCTTACAGAACCTGAAGACCGGCAAG 1725
|||
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
```

## RESULT 4

```
US-09-124-900-3
; Sequence 3, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: FREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124, 900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-124-900-3

Alignment Scores:
Pred. No.: 2,82e-207 Length: 1015
Score: 3475.00 Matches: 654
Percent Similarity: 93.72% Conservative: 33
Best Local Similarity: 89.22% Mismatches: 28
Query Match: 75.99% Indels: 18
DB: 3 Gaps: 4

US-09-610-313B-31 (1-2463) x US-09-124-900-3 (1-1015)
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGAAAGCGCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
268 -----CCGAGCGAGCAGAACCGCGCCCAACAGCCCGCCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
Qy 304 CGCGAGCTGCAGGTG-----CGGGCGCAACCCCGCAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
358 CAGGCGACCCCTG-----AACTTCCCGCACATCACCTGTGGCAGCGCGCCCTGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
412 ATCAAGTGGCGGCCAGATCAAGAGGCGCTGTGGACACCGCGCGCGCGACGACACCGTG 471
Db 81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100
472 CTGAGAGATGATGCTCGCCCGCAAGTGAAGCCCAAGATGATCGCGCGCGCATCGCGGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120
532 TTCTATCAAGTGGCGCGAGTACGACCATCTGATCGAGATCTGCGCAAGAGAGCGCATC 591
Db 121 PheLysValArgGlnTyrAspGlnIleLeuIleGluLeuLysGlyHisLysAlaIle 140
592 GGCACCGTGTGATCGGCGCCCAACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651
|||
```

```
QY 1726 TAGCCAGATGCGACCGCCACACCAACGACGTGAAGAGCTGACCGAGCGCGTGCAG 1785
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1786 AAGATCGCCATGGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCCGCTCCCATC 1845
Db 541 LysIleThrThrGluSerIleValIleTrpGlyThrProLysPheLysLeuProIle 560
QY 1846 CAGAAGGAGACTGGGAGACCTGTGTGGACCGACTACTGGCAGGGCCACCTGATCCCGAG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGluThrTrpGlnAlaThrTrpIleProGlu 580
QY 1906 TGGAGTTCTGTGAACACCCCGCTGTGAAGCTGTGTACCGACTGTGGAGAAGAGCC 1965
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTrpGlnLeuGluLysGluPro 600
QY 1966 ATCATCGCGCGGAGACCTTCTAGTGTGACCGCGCGGCGAGAGATCGTGAGCCTGACCGACCA 2025
Db 601 IleValGlyAlaGluThrPheTrpValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCCCGCTACGTGACCGACCGCGCGGCGGCGAGAGATCGTGAGCCTGACCGACCA 2085
Db 621 LysAlaGlyTrpValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2086 AACAGAGACCGAGCTGACGGCCATCCAGTGGCCCTGTGAGGACGCGGAGCGAGGTG 2145
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTrpLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATCGTGCAGACGCGAGTACGCGCTGGGATCATCCAGGCGCGCGCAAGAGC 2205
Db 661 AsnIleValThrAspSerGlnTrpAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2206 GAGACGAGCTGTGTGAACACGATCATCGACGAGCTGATCAAGAAGAGAGAGGTGTACTG 2265
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTrpLeu 700
QY 2266 AGTGGTGTGCGCGCCCAAGGGCATCGCGGCAACGAGCGAGATCGCAAGCTGGTGCAG 2325
Db 701 AlaTrpValProAlaHisLysGlyIleGlyLysAsnGlnValAspLysLeuValSer 720
QY 2326 AAGGCGATCGCAGGTGTCTTCTGCGGCGATCGAT 2364
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 5
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,028
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
```

```
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1015
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
; US-08-463-028-9

Alignment Scores:
Pred. No.: 2,82e-207 Length: 1015
Score: 3475.00 Matches: 654
Percent Similarity: 93.72% Conservatives: 33
Best Local Similarity: 89.22% Mismatches: 28
Query Match: 75.99% Indels: 18
DB: Gaps: 4

US-09-610-313B-31 (1-2463) x US-08-463-028-9 (1-1015)
QY 220 TTCTTCGCGAGACCTGGGCTTCCCCAGGGCAAGCGCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 268 -----CCCAGCGAGCAGAACCGCGCCACACCGCCACCCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGCGCACAACCCCGCAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCGACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGCGCCCTCGTGCAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAGGTGGCGCGCGCAGATCAAGAGCGCCCTGTGACACCGCGCGCGCGAGCACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTGCGCGCAACTGGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY 532 TTCATCAAGGTGCGCCAGTACGACACAGATCCTGTATCGAGATCTCGCGCAAGAAGCCCATC 591
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGCGCCCGCCCGTGAACATCATCGCGCGCGCGCATCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACTTCCCCATCAGCCCCATCAGACCGCGCGCTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```



QY 712 CCGGCATGACGCCCCCAAGGTGAAGCTGGCCCTGACCGAGGAGAAAGATCAAGGCC 771  
DB 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 GTGACGCCCATCTCTCGAGGAGATGGAGAGAGGCAAGATCAACAGATCGGCCCGAG 831  
DB 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACAAACACCCCGCTTCCGCCATCAAGAAAGACAGACCAAGATGCGGCAAG 891  
DB 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCGCGAGCTGAACAGCGCACCCAGGACTTCGTGGAGGTGCAGCTGGGC 951  
DB 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCCACCACCCCGCTTGAAGAAGAAGAGCGTGACCGTGTGGAGCGTGGCGAC 1011  
DB 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTACGTGCCCCCTGGAGGAGACTTCGGCAAGTACACCGCTTCACCATCCCC 1071  
DB 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACGAGACCCCGGCATCCGCTACCAAGTACAACGTGCTGCCCGCCAGGGCTGG 1131  
DB 301 SerIleAsnAsnGlnThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGCAGCCCGACATCTTCCAGAGCAGCATCAACCAAGATCTGGAGCCCTTCCGCGCC 1191  
DB 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCGAGATCGTACTACCG-----GCCCGCTGTAGTGGGCGAGGACCTG 1245  
DB 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
QY 1246 GAGATCGGCAGCACCGCCCAAGATCGAGAGCTGGCAAGCACCTGCTGCGCTGGGGC 1305  
DB 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuAspTrpGly 380  
QY 1306 TTCACACCCCGCAAGAACGACCAAGAGAGCCCGCTTCTGTGGATGGGCTACGAG 1365  
DB 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400  
QY 1366 CTGCACCCCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGAGCTGGACC 1425  
DB 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1426 GTGAACGACATCCAGAGCTGTGGGCAAGCTGAATGGGCCAGCAGATCTACCCCGGC 1485  
DB 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY 1486 ATCAAGTGGCCCATGTGTCAAGCTCTCGCGGCGCCCAAGGCCCTGACCGACATCGTG 1545  
DB 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1546 CCCTGTACCGAGGCGGAGCTGGAGCTGCCGAGAACCGCGAGATCCTGCGCGAGCCC 1605  
DB 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGlnAsnArgGluIleLeuLysGluPro 480  
QY 1606 GTGCACCGCGTGTACTACGACCCCGACAGACCTGGTGGCCGAGATCCAGAAAGAGGGC 1665  
DB 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1666 CACGACAGTGGACCTTACAGATCTTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725  
DB 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1726 TAGCCCAAGATGGCCCGCCACACCAAGCGTGAAGCTGACCGAGGCGGTGCAG 1785  
DB 521 TyrAlaAspMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1786 AAGATCGCCATGGAGATCGTGTATCTGGGGCAAGACCCCAAGTTCCGCGCTGCCCATC 1845

DB 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1846 CAGAAGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCCGAG 1905  
DB 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1906 TGGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGGTACCAGCTGGAGAGAGGAGGCC 1965  
DB 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600  
QY 1966 ATCATCGGCGCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCAGACCAAGATCGC 2025  
DB 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2026 AAGCGCGGTACTGTACCGACCGCGCGCGCGCAGAAATCGTGAAGCTGACCGACCCACC 2085  
DB 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2086 AACCAAGACCGAGCTGCAGGCGCATCCAGCTGCGCTGCGAGCACAGCGCGCAGAGTG 2145  
DB 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2146 AACATCGTGACCGACAGCCAGTACCGCTGGGCATCATCCAGCGCCCGACCCGCAAGAGC 2205  
DB 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2206 GAGAGCGAGCTGTGTGAACCGATCATCGAGCAGCTCATCAAGAGAGAGGTGTACTCG 2265  
DB 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2266 AGCTGTGTGCGCCCGCCCAAGGCGCATCGCGGCAAGCAGAGCAGATCGACAGCTGGTGAGC 2325  
DB 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2326 AAGGCGATCCGAAAGTGTCTTCTTGGAGCGCATCGAT 2364  
DB 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 6  
US-07-743-357-2  
; Sequence 2, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900

```

; TELFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: BH102
; US-07-743-357-2

Alignment Scores:
Pred. No.: 2,82e-207 Length: 1016
Score: 3475.00 Matches: 654
Percent Similarity: 93.72% Conservative: 33
Best Local Similarity: 89.22% Mismatches: 28
Query Match: 75.99% Indels: 18
DB: 2 Gaps: 4

US-09-610-313B-31 (1-2463) x US-07-743-357-2 (1-1016)

Qy 220 TTCTTCGCGAGGACCTGGCCCTTCCCGCAGGCAAGGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAepLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy 268 -----CCGAGCGAGCAGAACCGCGCCACAGCCCGCCACAGCCCGCCAGC 303
Db 21 ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg 40
Qy 304 CGCGAGCTGCAGGTG-----CGCGCGCACAACCCCGCAGCGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAenAenSerProSerGluAlaGlyAlaAspArg 60
Qy 358 CAGGCGCACCTG-----AACTTCCCGCAGATCACCTGTGGCAGCCCGCCCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAenPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
Qy 412 ATCAAGTGGCGCGCGAGATCAGGAGCGCTGTCGACACCGCGCGCGCGCGCGCGCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAenAspThrGlyAlaAspThrVal 100
Qy 472 CTGAGGAGATGAGCTCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
Qy 532 TTCATCAAGTGGCGCGAGTACGACCATCTGATCGAGATCTGCGCAAGAGCGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
Qy 592 GGCACCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
Db 141 GlyThrValLeuValGlyProThrProValAenIleIleGlyArgAenLeuLeuThrGln 160
Qy 652 CTGGGCTGCACTTCCCGCAAGTGAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
Db 161 IleGlyCysThrLeuAenPheProIleSerProIleGluThrValProValLysLeuLys 180
Qy 712 CCCGGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
Qy 772 CTGACCGCCATCTCGAGGAGATGAGAGAGGCGCAAGATCACCAGATCGCGCGCGCG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
Qy 832 AACCCCTTACACACCCCGCGTGTCCGCATCAAGAGAGAGAGAGAGAGAGAGAGAG 891
Db 221 AsnProIyrAenThrProValPheAlaIleLysLysLysAspSerThrLysIleTrpArgLys 240

```

```

Qy 892 CTGGTGGACTTCGCGAGCTGGAACAAGGCGACCCAGGACTTCTGGAGGTCGAGCTGGC 951
Db 241 LeuValAspPheArgGluLeuAenLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
Qy 952 ATCCCCCAGCCCGCGCGCTGGAAGAAGAAGAGCGTGCCTGCTGGAGCTGGCGCAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280
Qy 1012 GCCTACTTTCAGCGTGCCTTCGAGGAGACTTCGCAAGTACACCGCTTCACCATCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
Qy 1072 AGCATCAACAACGAGACCCCGCGCATCGCTACAGTACACGCTGCTGCCCGCGCTGG 1131
Db 301 SerIleAenAenGluThrProGlyIleArgTyrGlnTyrAenValLeuProGlnGlyTrp 320
Qy 1132 AAGGCGAGCCCGCAGCATCTTCCAGAGCAGATGACCAAGATCTCCGAGGCGCTTCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
Qy 1192 CGCAACCCCGAGATCGTGATCTTACCAG-----GCCCGCTGTACGTGGCGAGCAGCTG 1245
Db 341 GlnAenProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
Qy 1246 GAGATCGCGCAGACCCCGCGCAGATCGAGAGCTGCGCAAGACCTGCTCGCTGGCG 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
Qy 1306 TTCAACACCCCGCAGAGAGCAGACAGAGAGCGCCCTTCTGCTGATGGGTACGAG 1365
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTrpGlu 400
Qy 1366 CTGCACCCCGCAGAGTGGAGCCGTCAGCCCATCGAGTGCCTCCGAGAGAGAGAGTGGAC 1425
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
Qy 1426 GTGAACAGATCCAGAGCTGGTGGCAAGCTGAGTGAACCTGGCGCAGCAGATCTACCCGCG 1485
Db 421 ValAenAspIleGlnLysLeuValGlyLysLeuAenTrpAlaSerGlnIleTyrProGly 440
Qy 1486 ATCAAGTGGCGCAGCTGTGCAAGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCATG 1545
Db 441 IleLysValArgGlnLeuLysLysLeuLysLysLysLysLysLysLysLysLysLys 460
Qy 1546 CCCTTGACCGAGGAGCGCGAGCTGGAGCTGCGCGAGAACCGCGAGATCTCTCGCGAGCC 1605
Db 461 ProLeuThrGluGluAlaGluLeuLeuAlaGluAenArgGluIleLeuLysGluPro 480
Qy 1606 GTGACGCGCTGTACTACGACCCCGCAGAGCCTGCTGGCGCGAGATCCAGAGCAGCGC 1665
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
Qy 1666 CACGACGAGTGGACCTACAGATCTACAGAGCGCTTCAAGACCTGAAGACCGCGCAAG 1725
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysLysLysLysLysLysLysLysLys 520
Qy 1726 TAGCCCAAGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1785
Db 521 TyrAlaArgMetArgGlyAlaHisThrAenAspValLysGlnLeuThrGluAlaValGln 540
Qy 1786 AGATCGCCATGAGAGCAGATCGTGATCTGGGGCAAGACCCCGCAAGTTCGCGCTGCCCATC 1845
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysTrpLysPheLysLeuProIle 560
Qy 1846 CAGAGGAGACCTGGGAGACCTGTGACCGCTACTCTGGCAGCGCGCGCGCGCGCGCG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGluThrTrpGlnAlaThrTrpIleProGlu 580
Qy 1906 TGGAGTTCGTGAACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1965
Db 581 TrpGluPheValAenThrProProLeuValLysLeuTrpTrpGlnLeuGluLysGluPro 600
Qy 1966 ATCATCGCGCGCGAGACCTTCTTACGTGGAACCGCGCGCGCGCGCGCGCGCGCG 2025

```

```
Db 601 ILeValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCGCGGTACTGACCGACCGGGCGGCGAGAGATCGTAGCTGACCGTACCGAGACACC 2085
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2086 AACGAGAACCCAGCTGCGAGGCCATCCAGCTGGCGCTGCGAGACAGCGCAGCGAGGTG 2145
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATCGTGACCGACAGCCAGTAGTACGCCCTGGCGCATCATCCAGGCCCGCCGCAAGAGC 2205
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleleGlnAlaGlnProAspLysSer 680
QY 2206 GAGAGCGAGCTGTGAACCGATCATCGACGACGCTGATCAGAGAGAGGTGTACCTG 2265
Db 681 GluSerGluLeuValAsnGlnIleleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2266 AGCTGGGTGCCCGCCACAGGCGCATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGC 2325
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlnValAlaAspLysLeuValSer 720
QY 2326 AAGGCGATCCGCAAGTGCTGTTCCTCGAGCGCATCGAT 2364
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7
US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-309-572-17

Alignment Scores:
Pred. No.: 4,32e-207 Length: 1003
Score: 3472.00 Matches: 652
Percent Similarity: 95.15% Conservative: 34
Best Local Similarity: 90.43% Mismatches: 29
Query Match: 75.92% Indels: 6
DB: 4 Gaps: 3

US-09-610-313B-31 (1-2463) x US-09-309-572-17 (1-1003)
QY 220 TTCTTCGCGGAGGACCTGGCGCTTCCCGCAGGCGAAGCGCCGCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCAAACGCCCCACAGCGCGAGCTGCGAGTG- - - - -CGCGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCGCAGGCGACCCCTG- - - - -AACTTCCCGCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCGCTGTGTGAGCATCAAGTGGCGCGCGCGATCAAGGAGCGCCCTGCTG 447
```

```
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGACACACCGCTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGAGCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGGCGGATCGCGGCTTCATCAAGGTGCGCCAGGTACGACACAGATCCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGGCGCAAGAGCCCATCGCACCGCTGCTGATCGGCGCCACCGCCCGTGAACAC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTGAACCTTCCCATCAGCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTCCCGTGAAGCTGAAGCCCGGCATCGACGCGCCCAAGGTGAAGCAAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAGATGAGAGAGAGGCG 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGAGAACCCCTACAACACACCCCGCTGTCGCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGAGGTGCGAGCTGGGCATCCCGCACCCCGCGCGCTGAGAGAGAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer 260
QY 988 GTGACCGTCTGGAGCTGGCGCGACGCTACTTTCAGCGTGGCGCGCGCGAGCTTCCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGATCAACACAGAGACCCCGCGCATCCGCTACCAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAAGCTGCTGCCCGAGGCTGGAGAGTCGCGCAGCACCGCGCCCAAGATCGAGAGAGCTG 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320
QY 1168 AAGATCTGGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGATCTACACG- - - - -GCC 1221
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1222 CCCTCTAGCTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCCCAAGATCGAGAGAGCTG 1281
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGCTGGCTGGGCTTCCACCCCGCGCAAGAGAGCAGCAGAGAGAGAGGAGCCC 1341
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTCTCTGGATGGCTACGAGCTGCACCCCGCAAGTGGACCGCTGACCGCCATCGAG 1401
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCCGAGAGAGAGCTGGACCGTGAACGATCAACAGCATCGAAGCTGTGGCGCAAGCTGAAC 1461
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGCGCAGCGCATCTACCCCGCGCATCAAGTGGCGCGCGCTGCAAGCTGTGCGCGGC 1521
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
```

```
QY 1522 GCCAAGGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCGCGAGCTGGAGCTGGCCGAG 1581
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
441 ThrLysAlaLeuThrGluValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
QY 1582 AACCGCGAGATCCCTGGCGAGCCGCTGACCGCGCTGTACTACGACCCCGACGCAAGAGCTG 1641
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
461 AsnArgGluLeuLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1642 GTGGCCGAGATCCAGAAAGCGGCGACCGACAGATGGACCTACCATGATCTACGAGAGCCC 1701
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
481 IleAlaGluLeuGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
QY 1702 TTCAGAACTGAAGACCGGCAAGTACGCCAAGATGGCCACCGCCACCAACAGAGCTG 1761
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCGAGGCGGTGACAGATGCCATGCCATGAGAGCATCGTATCTGGGGCAAG 1821
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTyrGlyLys 540
QY 1822 ACCCCCAAGTTCCTGGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGACCGACTAC 1881
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
541 ThrProLysPheLysLeuProIleGlnLysGluThrTyrGluAlaTyrTyrThrGluTyr 560
QY 1882 TGGCAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1942 TGGTACCACTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTAGTGGAGCGGCC 2001
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGGCGCGCAGAAG 2061
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY 2062 ATCGTGAAGCTGACCGAGACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2121
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCAGACACCGCAGCGAGGTGAACATCGTGCACCGACAGCCAGTACGCCCTGGGCATC 2181
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCCAGGCCACCGCCGACAGAGCGAGCTGGTGAACCAAGATCATCGACGAGCTG 2241
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCAAC 2301
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
QY 2302 GAGCAGATCGACAGCTGGTGAAGAGGATCCGCAAGGTGTCTGCTGGAGCGGCATC 2361
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db |||
721 Asp 721
```

## RESULT 8

```
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR FILING DATE: 1998-11-26
```

```
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-718-096-17
```

## Alignment Scores:

Pred. No.:	4.32e-207	Length:	1003
Score:	3472.00	Matches:	652
Percent Similarity:	95.15%	Conservative:	34
Best Local Similarity:	90.43%	Mismatches:	29
Query Match:	75.92%	Indels:	6
DB:	4	Gaps:	3

US-09-610-313B-31 (1-2463) x US-09-718-096-17 (1-1003)

```
QY 220 TTCTTCCGCGAGGACCTGGCTTCCCGAGGGCAAGCCCGCGAGTTCCTCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCCCAACACGCCCCACCGCGAGCTGCAGGTG-----CGCGGCGACAAACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGGCGCGCGCGCGAGCGCGAGCGCACCCCTG-----AACTTCCCCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGGCGCCCTGGTGTGAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCCCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGAGCACCGCTGCTGAGGAGATGAGCTGCCCGCGCACTGGAGGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCGCGAGTACGACCATCTCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGGCGCAAGAGGCGCATCGGCACCGTGTGTGATCGGCCCGCCACCCCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGTGACCCAGCTGGGTGACACCTGCACTTCCCATCAGCCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCTGTCCTGTAAGCTGAAGCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGAGAGATCAAGAGCCCTGACCGCCCATCTGCGAGGAGATGGAGAGAGGCGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGAGAACCCCTACACACCCCGCTGCTGGCCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACACGACCAAGTGGCGCAAGCTGGTGAAGTTCCTCCGAGCTGAAACAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGAGGTGCAGCTGGGCATCCCCCACCGCGCGCTGAAAGAAAGAGAGC 987
```

Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysSer 260  
QY 988 GTGACCGTGTGACGTCGGCGAGCGCTACTTACGGTGGCCCTGACGAGGACTTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCGCTTACCATCCCGACGATCAACAACAGAGACCCCGCGCATCCGCTACCCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
QY 1108 TACAAGCTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGACGATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320  
QY 1168 AAGATCTGTGAGCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCGA-----GCC 1221  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
QY 1222 CCCTGTACGTGGCGAGCGACCTGGAGATCGGCAGCAGCCCGCCCAAGATCGAGGAGCTG 1281  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1282 CGCAAGCACCTGTGCTGGCGCTTCCACCACCCCGCAAGAAGCACCAGAGAGAGCC 1341  
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1342 CCTTCTGTGGATGGCTACGAGTGCACCCCGACCAAGTGGACCGTGAGCCCATCGAG 1401  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1402 CTGCCCGAGAGAGAGCTGGACCGTGAACAGCATCCAGAGCTGTGGCGAAGCTGAAAC 1461  
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1462 TGGCGCCAGCGAGATCTACCCCGCGCATCAAGTGGCGGAGCTGTGCAAGCTGTGCGCGC 1521  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
QY 1522 GCCAAGCCTGACCGACATCGTGCCCTGACCGAGAGCGCCGAGCTGGAGCTGGCGGAG 1581  
Db 441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuValAlaGlu 460  
QY 1582 AACCGGAGATCTGCGCGAGCCGCTGACGGGTGTACTACGACCCCGACAGGAGCCTG 1641  
Db 461 AsnArgGluIleLysGlnProValHisGlyValTyrAspProSerLysAspLeu 480  
QY 1642 GTGCGCGAGATCCAGAGCAGGCGCCACGACCTGGACCTTACGAGATCTACGAGGAGCC 1701  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500  
QY 1702 TTCAAGACCTGAAGCCGCAAGTAGCCCAAGATGGCACCGCCCGACCAACAGAGCTG 1761  
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520  
QY 1762 AAGCAGCTGACCGAGCGCGCAGAGATCGCATGGAGAGCATCGTATCTGGCGCAAG 1821  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
QY 1822 ACCCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTAC 1881  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTyr 560  
QY 1882 TGGCAGGCCACTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGGTGAAGCTG 1941  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
QY 1942 TGGTACAGCTGGAGAGGAGCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGGCC 2001  
Db 581 TrpTyrGlnLeuGlnLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600  
QY 2002 GCCAACCGCGAGACCAAGATCGGACGAGCGCGGTACTGTACCGACCGCGCGCGAGAG 2061

Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620  
QY 2062 ATCGTGAGCTGACCGAGACCCACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121  
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
QY 2122 CTGCGAGCAGCGCGCAGCGAGGTGAACATCGTGACCGAGCAGCAGTACGCCCTGGGCATC 2181  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
QY 2182 ATCCAGGCCCGCCGACCAAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTG 2241  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
QY 2242 ATCAAGAAGGAGAGGTGTACCTGAGTGGTGGCGCCGACCAAGGGCATCGCGCGCAAC 2301  
Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
QY 2302 GAGCATCGACCAAGCTGTGTGAGCAAGGCGATCCCGCAAGGTGTCTTCTGGACGCGCATC 2361  
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2362 GAT 2364  
Db 721 Asp 721  
RESULT 9  
US-07-743-357-5  
; Sequence 5, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1016 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: BRU

US-07-743-357-5

**Alignment Scores:**

Pred. No.:	6.64e-207	Length:	1016
Score:	3469.00	Matches:	652
Percent Similarity:	93.8%	Conservative:	36
Best Local Similarity:	88.9%	Mismatches:	27
Query Match:	75.8%	Indels:	18
DB:	2	Gaps:	4

US-09-610-313B-31 (1-2463) x US-07-743-357-5 (1-1016)

Qy	220	TTCTTCGCGAGAGACTGGCTTCCCCCAGGGCAAGGCCCGCGAGTTC-----	267
Db	1	PhePheArgGluAupLeuAlaPheLeuGlnGlyLysAlaAraGluPheSerSerGluGln	20
Qy	268	-----CCCAGCGAGCAGAACCGCGCCAAACAGGCCCCACACG	303
Db	21	ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg	40
Qy	304	CGCGAGCTGCGAGGTG-----CGGGCGGCACAAACCCCGCAGCGAGCGCGGGCCGACGGCC	357
Db	41	ArgGluLeuGlnValTrpGlyArgAspAenAenSerLeuSerGluAlaGlyAlaAspArg	60
Qy	358	CAGGGCACCCCTG-----AACTTCCCCAGATCACTGTGGCAGCGCCCTGTGTGAGC	411
Db	61	GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr	80
Qy	412	ATCAAGGTGGCGCGCCAGATCAAGGAGGCCCTGCTGGACACGGCGCGCAGACACCCCTG	471
Db	81	IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal	100
Qy	472	CTGAGCAGAGATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC	531
Db	101	LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly	120
Qy	532	TTCATCAAGGTGCGCCAGTACGACACAGATCTCTGATCGAGATCTGCGGCAAGAAGCCATC	591
Db	121	PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle	140
Qy	592	GGCACCGTGTGATCGGCGCCACACCCCGTGAACATCATCGCGCGCGCAACATGCTGACC	651
Db	141	GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln	160
Qy	652	CTGGCGTGCACCTGAACCTCCCATCATGCCCATCGAGACCGTSCCGTGAAGCTGAAG	711
Db	161	IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys	180
Qy	712	CCCGGCATGGACGCGCCCAAGGTGAAGCAGTGGCCCTGTACCGGAGGAGAAGATCAAG	771
Db	181	ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluLysIleLysAla	200
Qy	772	CTGACCGCCATCTCGAGGAGATGGAGNAGGAGGGCAGATCACCMAGATCGGCCCGAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu	220
Qy	832	AACCCCTACACACCCCGTGTTCGCCATCAAGAAGAGGACAGCACCAGTGGCGGCAAG	891
Db	221	AsnProLysAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys	240
Qy	892	CTGTTGAGCTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGC	951
Db	241	LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCCACCCCGCGGCTGAAGNAGNAGAGAGCGTGCACCGTGTGAGCTGGCGCAG	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTPACTTTCAGCGTCCCTCGACGAGGACTTCGCGAGTAGTACACCGCCTTCACCAT	1071
Db	281	AlaTyrPheSerValProLeuAspGluAspPheArgLysIleThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACAACAGAGACCCCGGATCCGCTACCAGTACAACAGTGTCTCCCGCAGGCTGG	1131

Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnProAspLysSer 680  
QY 2206 GAGCGCGAGTGTGACAGATCATCGACAGCTGATCAAGAGGAGAGTGTACCTG 2265  
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGluLysValTyrLeu 700  
QY 2266 AGCTGGGTGCGCGCCCAAGGCGATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGC 2325  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2326 AAGGCGATCCGCAAGTGTCTTCTGGACGCGATCGAT 2364  
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733  
RESULT 10  
US-07-743-357-3  
; Sequence 3, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1016 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: BH5  
; US-07-743-357-3  
Alignment Scores:  
Pred. No.: 7.66e-207 Length: 1016  
Score: 3468.00 Matches: 652  
Percent Similarity: 93.72% Conservative: 35  
Best Local Similarity: 88.95% Mismatches: 28  
Query Match: 75.84% Indels: 18  
DB: 2 Gaps: 4  
US-09-610-313B-31 (1-2463) x US-07-743-357-3 (1-1016)

QY 220 TTCTTCCGCGAGGACCTGCCTTCCCTCCAGGCAAGCGCCCGCAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerGluGln 20  
268 -----CCACGCGAGCAGAACCGCGCCCAACAGCCCCCACCAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
QY 304 CGCGAGCTCCAGGTG-----CGCGCGACAAACCCCGCAGCGAGCGCGCGCGAGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGCGACCTCG-----AACTTCCCGCAGATCACCTGTGCGAGCGCGCCCTGTGTGAGC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGTGGCGCGCGCAGATCAAGAGGCGCTGTGTGGACACCGCGCGCGCAGCACCGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
QY 472 CTGGAGGATGAGCTGCGCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531  
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120  
QY 532 TTCATCAAGTGGCGCGCGCAGTACGACACGATCTCGATCGAGATCTGCGCGCAAGAGGCCATC 591  
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGlyLysGlyHisAlaIle 140  
QY 592 GGCACCGTGTGATCG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
QY 652 CTGGCGTGCACCTGAACTTCCCATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGCGCATGAGCG 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCGCATCTGCGAGGAGATGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTTACAACACCCCGCTGTCGCATCAAGAAGAAGCAGACCAAGTGGCGCGCAAG 891  
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGGTGGACTTCGCGAGCTGAAACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGCG 951  
Db 241 LeuValAspPheArgGluLeuAsnArgA-gThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCGCACCCCG 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTTCAGCGTGCCTCGGACGAGGACTTCCGCAAGGTACACCGCTTCCACATCCCC 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTrpThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACAACGAGACCCCGCGCATCCGCTACCATCAACAGTGTGCGCGCGCGCGTGG 1131  
Db 301 SerIleAsnAsnGluThrProGlySerGlyTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGCGAGCG 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
QY 1192 CGCAACCCCGAGATCGTGTACTACCAAG-----GCCCGCTGTGTGTGTGTGTGTGTGT 1245  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
QY 1246 GAGATCG 1305







ADDRESSEE: KIRBY EADES GALE BAKER  
STREET: Box 3432, Station D  
CITY: Ottawa  
STATE: Ontario  
COUNTRY: Canada  
ZIP: K1M 1H8  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.3.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/743,357  
FILING DATE: 21-AUG-1991  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/CA90/00062  
FILING DATE: 23-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Gale, Edwin J.  
REGISTRATION NUMBER: 28,584  
REFERENCE/DOCKET NUMBER: 30924-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (613) 237-6900  
TELEFAX: (613) 237-0045  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1003 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHEITICAL: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Human immunodeficiency virus type 1  
STRAIN: ELI  
US-07-743-357-10

Alignment Scores:			
Pred. No.:	2.56e-206	Length:	1003
Score:	3459.50	Matches:	646
Percent Similarity:	95.28%	Conservative:	40
Best Local Similarity:	89.72%	Mismatches:	29
Query Match:	75.65%	Indels:	5
DB:	2	Gaps:	3
US-09-610-313B-31 (1-2463) x US-07-743-357-10 (1-1003)			
Qy	220	TTCTTCGGCAGACCTGGCTTCCCCCAGGCGCAAGGCCGCGAGTTCCTCCCGAGCGAGCAG	279
Db	1	PhePheArgGluAsnLeuAlaPheProGlnGlyIysAlaGlyGluLeuSerProIysGln	20
Qy	280	AACCGCGCCACACGCCACACGCGCGCAGCTGTCAGGTGCGCGGC--GACCAACCCCGC	336
Db	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu	40
Qy	337	AGCGAGCGCGCGCGCGAGCGCGCAGGCGACCCCTG-----AACTTCCCCAGATCACCCCTG	390
Db	41	SerLysThrGlyAlaGluArgGlnGlyThrValSerPheAsnPheProGlnIleThrLeu	60
Qy	391	TGGCAGCGCCCCCTGGTGATCATAGTGGGCGCGCCAGATCAAGAGGCCCCCTGCTGCAC	450
Db	61	TrpGlnArgProLeuValAlaIleIysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp	80
Qy	451	ACCGCGCGCCAGCAGCACCGTGCTGGAGAGATGAGCCTGCCCGCGCAAGTGGGAAGGCCCAAG	510
Db	81	ThrGlyAlaAspAAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys	100
Qy	511	ATGATCGGCGGCATCGCGCGCTTCATCAAGTGGCGCCAGTACGACACAGATCCTGATCGAG	570
Db	101	MetIleGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProIleGlu	120

Qy	571	ATCTGGCGCAAGAGCCGATCGGCATCGGTGCTGATCGGCCCAACCCCGGTGAACATCATC	630
Db	121	IleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle	140
Qy	631	GGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACCTTCCCCATCAGCCCCATCGAG	690
Db	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
Qy	691	ACCGTGCCCGTGAAGCCCGGCATCGACGGCCCAACAGGTGAAGCAGTGGCCCTTG	750
Db	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
Qy	751	ACGAGGAGAAGATCAAGGCCCTGACGGCCATCTCGAGGAGATGAGAAGAGGCGCAAG	810
Db	181	ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys	200
Qy	811	ATCACCAAGATCGGCCCGCAGAACCCCTCAACACACCCCGTGTTCGCCATCAAGAGAAGA	870
Db	201	IleSerArgIleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys	220
Qy	871	GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC	930
Db	221	AspSerThrLysIleArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
Qy	931	TTCTGGAGGTGAGCTGGGCATCCCCACCCCGCCGCGCTGAAGAAGAAGAGACGCTG	990
Db	241	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	260
Qy	991	ACCGTGTGAGCTGGCGAGCGCTACTTTCAGGTGCGCCCTGGACGAGACTTCCGCAAG	1050
Db	261	ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys	280
Qy	1051	TACACCGCTTACCATCCCGAGCATCAACAACGAGACCCCGGCATCCGCTACCAAGTAC	1110
Db	281	TyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr	300
Qy	1111	AACGTGTGCCCCAGGGCTGGAAGGGCAGCCCGCAGCATTTTCAGAGCAGCATGACCAAG	1170
Db	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
Qy	1171	ATCCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCTTACCAG-----GCCCCC	1224
Db	321	IleLeuGluProPheArgLysGlnAsnProGluMetValIleTyrGlnTyrMetAspAsp	340
Qy	1225	CTGTAGCTGGCAGCACCTCGAGATCGGCAGCACCGCGCCCAAGATCGAGAGCTGCGC	1284
Db	341	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg	360
Qy	1285	AAGCACTGCTGGCTGGGGCTTACACACCCCGCAAGAAGCACAGAGAGAGAGCCCCC	1344
Db	361	GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro	380
Qy	1345	TTCCCTGTGATGGCTACAGCTGCACCCCGACAGTGCACCGTGCAGGCCATCGAGCTG	1404
Db	381	PheLeuTrpMetGlyTyrGluLeuHisProAspLysIleTrpThrValGlnSerIleLysLeu	400
Qy	1405	CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCGAGCTGAAGTGG	1464
Db	401	ProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGluArgLeuAsnTrp	420
Qy	1465	GCCAGCCAGATTCACCCCGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCC	1524
Db	421	AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr	440
Qy	1525	AAGGCCCTGACCGACATCTGCCCCGTGACCGAGGAGCCGAGCTGGAGCTGGCCGAGAC	1584
Db	441	LysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGluAsn	460
Qy	1585	CGGAGATCTCGCGCAGCCCGTGCACGCGGTGTACTACGACCCCAAGCAGGACCTGGTG	1644
Db	461	ArgGluIleLeuLysGluProValHisGlyValTyrTrpProSerLysAspLeuIle	480
Qy	1645	GCCGAGATCCAGAAGCAGGGGCCACGACAGTGGACCTTACAGATCTTACAGAGAGCCCTTC	1704

```
Db 481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrglnIleTyrglnGluProPhe 500
QY 1705 AGAACTGAAGACCGGCAAGTAGCAGCAAGATCGGCACCGCCACACCAACGACGTGAAG 1764
Db 501 LysAsnLeuLysThrGlyLysTyraAlaArgMetArgGlyAlaHisThrAsnAspValLys 520
QY 1765 CAGCTGACGAGCGCGTGCAGAGATCGCCATCGAGAGCATCGTGTGGGGCAAGACC 1824
Db 521 GlnLeuAlaGluAlaValGlnArGileSerThrGluSerIleValIleTrpGlyArgThr 540
QY 1825 CCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGGTGGACCTACTG 1884
Db 541 ProLysPheArgLeuProLysGlnLysGluThrTrpGluThrTrpAlaGluTyTrp 560
QY 1885 CAGGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCCCTGGTGAAGCTGTGG 1944
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580
QY 1945 TACGAGCTGGAGAGGAGCCATCATCGCGCGGAGACCTTCTACGTGGACGGCGCGCC 2004
Db 581 TyrGlnLeuGlnLysGluProIleIleGlyAlaGluThrPheTyValAspGlyAlaAla 600
QY 2005 AACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGCGCGCGCAGAAATC 2064
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyValThrAspArgGlyArgGlnLysVal 620
QY 2065 GTGAGCTGACGAGACCAACCAACGAGACCGAGCTGCGAGCCATCCAGTGGCGCTG 2124
Db 621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
QY 2125 CAGGACGCGGACGAGGTGACATCGTACCGACGACGAGTACCGCTGGGCGCATCATC 2184
Db 641 GlnAspSerGlyLeuGluValAlaIleValThrAspSerGlnTyValAlaLeuGlyIle 660
QY 2185 CAGCGCCAGCCGACAGAGGAGCGAGCTGGTGAACCAAGATCATCGAGCGCTGATC 2244
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
QY 2245 AGAAGAGAAAGTGTAACCTGAGCTGGTGGTCCCGCCCAAGGGCATCGCGGCAAGGAG 2304
Db 681 LysLysGlnLysValTyLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700
QY 2305 CAGATGACAGAGCTGGTGACAGGCGATCCGCAAGGTGCTGTTCTGGACGGCATCGAT 2364
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
```

## RESULT 13

```
US-07-743-357-8
; Sequence 8, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J. 28,584
; REGISTRATION NUMBER: 30924-2
; REFERENCE/DOCKET NUMBER: 30924-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: RF
; US-07-743-357-8
```

## Alignment Scores:

Pred. No.:	3,936-206	Length:	1003
Score:	3456.50	Matches:	647
Percent Similarity:	95.28%	Conservative:	39
Best Local Similarity:	89.86%	Mismatches:	29
Query Match:	75.58%	Indels:	5
DB:	2	Gaps:	3

US-09-610-313B-31 (1-2463) x US-07-743-357-8 (1-1003)

```
QY 220 TTCTTCGGGAGGACCTGGCTTCCCGAGGCGCGCGAGTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGlnLeuSerSerGluGln 20
QY 280 AACCGCGCCCAACAGCCCGACCGCGAGCTGCGAGTGGCGCGC---GACAAACCCCGC 336
Db 21 ThrArgAlaAsnSerProThrArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40
QY 337 AGCGAGCGCGCGCGAGCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
Db 41 SerGluAlaGlyLysAspArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60
QY 391 TGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450
Db 61 TrpGlnArgProIleValThrValLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80
QY 451 ACCGCGCGCGCGAGCACACCGCTGCTGGAGGAGATGAGCTGCCCGCGCAAGTGAAGCCCAAG 510
Db 81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100
QY 511 ATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
Db 101 MetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuLeuGlu 120
QY 571 ATCTGGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
Db 121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140
QY 631 GGCGCGCAACATGCTGACCCAGCTGGCGCTGCACCTGCACTTCCCGCGCGCGCGCGCGCG 690
Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160
QY 691 ACCGTCCCGCGTGAAGTGAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180
QY 751 ACCGAGGAGAGAGTGAAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810
Db 181 ThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLys 200
```

QY 811 ATCCAGAGTCCGCGCCGAGAACCCCTACACACCCCGTGTTCGCCATCAAGAGAG 870  
Db 201 IleserlysfleGlyProGluuAsnProTyrAsnThrProValPheAlaIleLysLys 220  
QY 871 GACAGACCAAGTGGCGCAAGCTGGTACTTCGCGAGCTGAACAAGCGCACCGAGAC 930  
Db 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240  
QY 931 TTCTGGAGGTGACGTGGGATCCCCACCCCGCGCTGGAAGAAGAAGAGCGTG 990  
Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysVal 260  
QY 991 ACCGTGTGACGTGGCGGAGCGCTACTTCAGCTGCGCCCTGGAGGAGCTTCGCAAG 1050  
Db 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysGluPheArgLys 280  
QY 1051 TACACCGCTTACCATCCCGAGCATCAACAACAGACCCCGCATCCGCTACCATGAC 1110  
Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgIleArgTyrGlnTyr 300  
QY 1111 AACGTGTGCTCCCGAGGCTGGAAGGCGAGCCCGCATCTTCCAGAGCAGCATGACCAAG 1170  
Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
QY 1171 ATCTGTGAGCCTTCCGCGCCGCGCAACCCCGAGATCGTATCTACAG-----GCCGCC 1224  
Db 321 IleLeuGluProPheLysLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1225 CTGTACCTGGCGAGCATGGAGATCGGCAGACCGCGCAAGATCGAGGACTCGCG 1284  
Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLysIleGluGluLeuArg 360  
QY 1285 AAGCACCTGTGCTGGGCTTCCACCAACCCCGCAAGAGCAGCAAGAGGAGCCGCC 1344  
Db 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
QY 1345 TTCTGTGGATGGCTACGAGCTCACCCCGCAAGTGGACCGTGGAGCTGGCCGAGAC 1404  
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeu 400  
QY 1405 CCGGAGAGGAGACTCGACCGTGAACACATCCAGAGCTGGTGGCAAGCTGAACCTGG 1464  
Db 401 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
QY 1465 GCCAGCAGATCTACCCCGCGCATCAAGTGGCGGAGCTGTGCAAGCTGCTGCGCGGCC 1524  
Db 421 AlaserGlnIleTyrAlaGlyIleLysValLysGlnLeuCyLysLeuLeuArgGlyThr 440  
QY 1525 AAGCCCTGACGACATCGTGCCTCCCTGACGAGAGCGCGAGCTGGAGCTGGCCGAGAC 1584  
Db 441 LysAlaLeuThrGluValValGlnLeuThrLysGluAlaGluLeuGluAlaGluAsn 460  
QY 1585 CGGAGATCTGCGCGAGCGCTGACGCGGTGTACTACGACCCCGCAAGGAGCTGGTG 1644  
Db 461 ArgGluIleLysGluProValHisGlyValTyrTrpAspProSerLysAspLeuIle 480  
QY 1645 GCCAGATCCAGAGCAGGCGCACGACAGTGGACCTTACAGATCTACGAGGAGCCCTTC 1704  
Db 481 AlaGluIleGlnLysGlnGlyGlnTrpThrTrpThrGlnIleTyrGlnLysProPhe 500  
QY 1705 AAGAACTGAAGCCGCAAGTACGCCAAGATGGCACCGCCCAACACACAGAGTGAAG 1764  
Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520  
QY 1765 CAGCTGACCGAGCGCTGCAGAGATCGCCATGAGAGCATCGTATCTGGGGCAAGAC 1824  
Db 521 GlnLeuThrGluAlaValGlnLysValAlaThrGluSerIleValIleTrpGlyLysThr 540  
QY 1825 CCCAAGTTCGCGCTCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACGACTACTGG 1884  
Db 541 ProlysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTyrTrp 560  
QY 1885 CAGGCCACCTGGATCCCGGAGTGGGAGTTCGTGTGAACACCCCGCCCTGGTGAAGCTGTGG 1944

Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
QY 1945 TACCAGCTGGAGAGGAGCCCATCATCGCGCCCGAGACCTTCTACGTGGAGCGCGCCGCC 2004  
Db 581 TyrGlnLeuGluLysGluProIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 2005 AACCGCAGAGCAAGATCGGCAAGCGCGGTACTGTGACCGAGCCGCGCGCGCAGAGATC 2064  
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
QY 2065 GTGAGCTGACGAGACCAACCAAGACCGAGTGCAGGCGCATCCACCTGCGCCCTG 2124  
Db 621 ValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2125 CAGCAGACCGCGCAGGAGTGAAATCATCGTACCCACAGCAGCAGTACGCGCTCGGCATCATC 2184  
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
QY 2185 CAGCGCCAGCCGCAAGAGCAGGAGCGAGCTGTGTGAACAGATCATCGAGCAGCTGATC 2244  
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680  
QY 2245 AAGAGAGAGAGCTGTACTGAGCTGGTGGTCCCGCCACAGGCGCATCGCGGCAACGAG 2304  
Db 681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
QY 2305 CAGATCCACAGCTGTGTGAGCAGGCGCATCGCAAGGTGTCTTCTGGAGCGGCATCGAT 2364  
Db 701 GlnValAspArgLeuValSerThrGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 14  
US-07-743-357-7  
; Sequence 7, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6500  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1004 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein

;  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: SF2  
; US-07-743-357-7

Alignment Scores:  
Pred. No.: 1,15e-205 Length: 1004  
Score: 3449.00 Matches: 646  
Percent Similarity: 94.87% Conservative: 38  
Best Local Similarity: 89.60% Mismatches: 31  
Query Match: 75.42% Indels: 6  
DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x US-07-743-357-7 (1-1004)

QY 220 TTCTTCGGGAGGAGCTGGCTTCCCGCCAGGGAGCGCGCGAGTTCCTCCAGCGAGCAG 279  
DB 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 280 AACCGGCGCAACAGCCCGACCGCGGAGCTCGAGTGGCGCGC-----GACACCCCG 333  
DB 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40  
QY 334 CGCAGCGAGCGCGCGCGAGCGCCAGGCGCACCCCTG-----AACTTCCCGCCAGATCAC 387  
DB 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
QY 388 CTGTGGCAGCGCCCTGTGTGATCATCAAGTGGCGGCGCAGATCAAGGAGCGCTGCTG 447  
DB 61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80  
QY 448 GACACCGCGCGCGAGCACCGTGTGTGAGGAGATGAGCTGCGCGCGAGTGGAGCGCC 507  
DB 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100  
QY 508 AGATGATCGCGCGCGCTGCGCGCTTCATCAAGTGGCGCGAGTACGACCGAGTCTGTGATC 567  
DB 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProVal 120  
QY 568 GAGATCTGCGCGCAAGAGCGCATCGGCACCGTCTGTATCGCGCGCCACCGCGCGTGAACATC 627  
DB 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
QY 628 ATCGCGCGCAACATGTCAGCCAGCTGGGTGTGACCTGACCTGACCTTCCCATCAGCCCATC 687  
DB 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
QY 688 GAGACCGTGCCTGAGCTGAAGCGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCG 747  
DB 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGGAGGAGATCAAGCGCGCTGACCGCGCATCTGCGAGGAGATGGAAGAGGAGGC 807  
DB 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AGATCACCAAGATCGCGCGCGCGAGAACCCCTACAAACCGCGCGTGTGCGCATCAAGAG 867  
DB 201 LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 220  
QY 868 AAGCAGACCAAGTGGCGCGCAAGCTGTGGACTTCGCGAGCTGAACAGCGCGCACCCAG 927  
DB 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
QY 928 GACTTCTGGGAGGTGCGAGCTGGGCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 987  
DB 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
QY 988 GTGACCGTGTGAGTGGCGCGAGCGCTACTTACGTGGTGGCGCGCGCGCGCGCGCGCG 1047  
DB 261 ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg 280

QY 1048 AAGTACACCGCCTTACCATCCCGCATCAACACGAGACCGCGCGCATCCGCTACCAG 1107  
DB 281 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 300  
QY 1108 TACAAGCTGTCTCCCGAGCGGTGGAAGGCGACCGCCAGCATCTTCAGAGCGAGCATGACC 1167  
DB 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTCTGGAGCCCTTCCCGCGCGCAACCGCGAGATCGTGTATCTACAG-----GCC 1221  
DB 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyGlnTyMetAsp 340  
QY 1222 CCCCTGTACGTGGCGAGCGACCTGCGAGATCGCGCGACCGCGCGCGCGCGCGAGAGCTG 1281  
DB 341 AspLeuTyThrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1282 CGCAGACCTGTCTGGCGCTTCCCGCGCGTTCACACCGCGCGCAACGAGACCGAGAGAGCC 1341  
DB 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1342 CCCTTCTGTGTGGCTACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1401  
DB 381 ProPheLeuTrpMetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleMet 400  
QY 1402 CTGCGCGAGAGGAGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1461  
DB 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1462 TGGCGCGAGCGAGATCAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1521  
DB 421 TrpAlaSerGlnIleTyAlaGlyIleLysValLysGlnLysLeuLysLeuArgGly 440  
QY 1522 GCCAAGCG 1581  
DB 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGlu 460  
QY 1582 AACCGCGAGATCTCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1641  
DB 461 AsnArgGluIleLeuLysGluProValHisGluValTyThrAspProSerLysAspLeu 480  
QY 1642 GTGCGCGAGATCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1701  
DB 481 ValAlaGluIleGlnLysGlnGlyGlnTyThrTrpThrTyGlnIleTyGlnGluPro 500  
QY 1702 TTCAGAACTCAAGACCG 1761  
DB 501 PheLysAsnLeuLysThrGlyLysTyAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
QY 1762 AAGCAGCTGACCGAGCG 1821  
DB 521 LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys 540  
QY 1822 ACCCCCAAGTTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1881  
DB 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpMetGluTy 560  
QY 1882 TGGCG 1941  
DB 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580  
QY 1942 TGTACACCTGAGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2001  
DB 581 TrpTyThrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyValAspGlyAla 600  
QY 2002 GCCAACCGCGAGACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2061  
DB 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyValThrAspArgGlyArgGlnLys 620  
QY 2062 ATCGTGTGAGCTGACCG 2121  
DB 621 ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
QY 2122 CTGCGAGCAGCG 2181

```

Db      641 LeuGlnaspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy      2182 ATCCAGGCCCCAGCCGCAAGAGCGAGAGCGAGCTGGTGAACACGAGATCATCGAGCAGCTG 2241
Db      661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy      2242 ATCAAGAGGAGAGAGGTGTACTGAGCTGGGTGGCCCGCCCAAGAGGCGATCGGGCGCAAC 2301
Db      681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
Qy      2302 GAGCAGATCGACAGCTGGTGAAGGCGATCGGAGGTGCTTCCTGGAGCGGCATC 2361
Db      701 GlnValAlaAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720
Qy      2362 GAT 2364
Db      721 Asp 721

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALEB- INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAULIERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSI, Francoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/09/319,588C
; CURRENT FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

Alignment Scores:
Pred. No.:      3,17e-189      Length:      1014
Score:          3183.50      Matches:      583
Percent Similarity: 90.25%      Conservative: 74
Best Local Similarity: 80.08%      Mismatches: 58
Query Match:      69.62%      Indels:      13
DB:              4          Gaps:      4

US-09-610-313B-31 (1-2463) x US-09-319-588C-6 (1-1014)
Qy      220 TTCTTCGCGAGGACCTCGGCTTCCCGCGGCAAGCGCCGAGTTCCTCCGAGCGAGCAG 279
Db      1 PhePheargGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProAspAsn 20
Qy      280 AAC-----CGCGCAACAGCCCAACCGAGCGCGGAGTGCAGGTG----- 318
Db      21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40
Qy      319 ---CGCGGCGACACCCCGCAGCGAGCGCGCGCGCGCGCCAG-----GGCACC 366
Db      41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60
Qy      367 CTGAACCTCCCGCAGATCACCCTGTGGCAGCGCCCTGTGGTGAACATCAAGGTGGCGGC 426
Db      61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
Qy      427 CAGATCAAGGAGGCGCTGTGGACACCGCGCGCGCGACACCGTGTCTGGAGGAGATGAGC 486

```

```

Db      81 GluValargGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluLeuGln 100
Qy      487 CTGCCCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCAAGGTGGCG 546
Db      101 LeuGluGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValarg 120
Qy      547 CAGTACACGACGATCTCGATCGAGATCTGCGGCAAGAGGCCATCGGCACCGTGTGTATC 606
Db      121 GlnTyrAspAsnIleThrValAspIleGlnArgLysAlaValGlyThrValLeuVal 140
Qy      607 GGCCCGCAGCCCGCGACATCATCGCGGCAACATGCTGACCCAGCTGGGTGGCGACCCGTG 666
Db      141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160
Qy      667 AACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGGAAGCCCGCGCATGACCGGC 726
Db      161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180
Qy      727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGAGCCCTGACCGCATCTGC 786
Db      181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200
Qy      787 GAGGAGATGAGAGAGGAGGCGCAGATCACCAAGATCGCGCCCGAGAACCCCTACACACC 846
Db      201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220
Qy      847 CCGGTGTTCCCATCAAGAAGAAGAGCAGCAGCAAGTGGCGCAAGCTGGTGGACTTCCGC 906
Db      221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhearg 240
Qy      907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGGCATCCCCACCCCGCC 966
Db      241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260
Qy      967 GGCTGAAGAAGAAGAGGCGTGCCTGCTGCGAGTGGCGCGCGCTACTTCTAGCGTG 1026
Db      261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280
Qy      1027 CCGCTGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAACAACGAG 1086
Db      281 ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu 300
Qy      1087 ACCCGCGCATCCGCTACAGTACACGCTGCTGCCAGGCTGGAAGGCGAGCCCGCAGC 1146
Db      301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320
Qy      1147 ATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCCGAGTC 1206
Db      321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340
Qy      1207 GTGATCTACCAG-----GCCCGCTGTACTGGGCGAGCGACCTGGAGATCGCGCAGCAC 1260
Db      341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360
Qy      1261 CGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGTGCGCTGGGGCTTCACACCCCGAC 1320
Db      361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380
Qy      1321 AAGAGCACCAGAGAGGAGCCCGCTTCTGTGGATGGGTACGAGCTGCACCCCGCAGCAG 1380
Db      381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400
Qy      1381 TGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAG 1440
Db      401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420
Qy      1441 AAGTGTGTGGCGCAAGCTGAACCTGGCGCAGCAGATCTACCCCGCATCAAGGTGGCGCAG 1500
Db      421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440
Qy      1501 CTGTGCAAGCTGTGCGCGCGCCCAAGCGCTGACCGACATCGTGGCCCTGACCGAGGAG 1560

```



```

Db      441  LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460
QY      1561  GCCGAGCTGGAGCTGGCCGAGAACCGGAGATCCTGCGCGAGCCCTGTGCACGCGGTGTAC 1620
Db      461  AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480
QY      1621  TACGACCCCAAGCAAGACCTGGTGGCCGAGATCCAGAAGCAGGCGCCACCAAGTGGACC 1680
Db      481  TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnTyrThr 500
QY      1681  TACGAGATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Db      501  TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520
QY      1741  ACCGCCCCACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAG 1800
Db      521  SerAlaHisThrAsnAspIleLysGlnLeuValGluValArgLysValAlaThrGlu 540
QY      1801  AGCATCGTGATCTGGGCGCAAGACCCCAAGTTCGGCTGCCCATCCAGAGGAGACCTGG 1860
Db      541  SerIleValIleTyrGlyLysThrProLysPheArgLeuProValGlnLysGluValTyr 560
QY      1861  GAGACCTGGTGACCGCACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1920
Db      561  GluAlaTyrTrpThrAspHisTrpGlnAlaThrTrpIleProGluTyrGluPheValAsn 580
QY      1921  ACCCCGCCCTGTGAGCTGTGTACCAAGTGGAGAGAGAGCCCATCATCGGCGCGGAG 1980
Db      581  ThrProLeuValLysLeuTyrTrpGlnLeuGluThrGluProLysSerGlyAlaGlu 600
QY      1981  ACCTTCTTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2040
Db      601  ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620
QY      2041  ACCGACCGGCGCGCGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db      621  ThrAspArgGlyArgGlnLysValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640
QY      2101  CTCGAGGCCATCCAGCTGGCGCTGCAGGACAGCGGAGGAGTGAACATCGTGACCGAC 2160
Db      641  LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660
QY      2161  AGCCAGTACGCCCTGGGCATCATCCAGGCCCGCAGCCAGCAAGAGCGAGCGAGCTGGTG 2220
Db      661  SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680
QY      2221  AACGAGATCATCGAGAGCTGATCAGAAGGAGAAGTGTACCTGAGCTGGTGGCCGCC 2280
Db      681  SerGlnIleIleGluGluLeuIleLysLysLysGluArgValTyrLeuSerTrpValProAla 700
QY      2281  CACAAGGGCATCGCGGCAACGAGCAGATCGCAAGCTGGTGAGCAAGGGCATCCGCAAG 2340
Db      701  HisLysGlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720
QY      2341  GTCTGTCTCTGGACGGCATCGAT 2364
Db      721  IleLeuPheLeuAspGlyIleGlu 728

```

Search completed: June 2, 2005, 05:12:37  
Job time : 109 secs

GenCoré version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 03:13:23 ; Search time 33.9172 Seconds  
(without alignments)  
10815.332 Million cell updates/sec

Title: US-09-610-313B-32  
Perfect score: 4555  
Sequence: 1 gtcagccaccatggccga.....gggtagcaccggtgaattc 2457

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09610313/runat\_31052005\_155137\_15173/app\_query.fasta\_1.7893  
-DB-Issued\_Patents\_AA -QPMF-fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1 101 @runat\_31052005\_155137\_15173 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6  
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3454	75.8	1005	2	US-07-743-357-1
2	3450.5	75.8	1003	2	US-07-743-357-9
3	3432	75.3	1015	3	US-08-463-210-9
4	3432	75.3	1015	3	US-09-124-900-3
5	3432	75.3	1015	4	US-08-463-028-9
6	3432	75.3	1016	2	US-07-743-357-2
7	3429	75.3	1003	4	US-09-309-572-17
8	3429	75.3	1003	4	US-09-718-096-17
9	3426	75.2	1016	2	US-07-743-357-5
10	3425	75.2	1016	2	US-07-743-357-3
11	3422	75.1	1016	2	US-07-743-357-4
12	3416.5	75.0	1003	2	US-07-743-357-10

13	3413.5	74.9	1003	2	US-07-743-357-8	Sequence 8, Appli
14	3406	74.8	1004	2	US-07-743-357-7	Sequence 7, Appli
15	3140.5	68.9	1014	4	US-09-319-588C-6	Sequence 6, Appli
16	3098	68.0	913	2	US-07-743-357-22	Sequence 22, Appl
17	3076.5	67.5	1350	4	US-09-952-060-35	Sequence 35, Appl
18	3035	66.6	913	2	US-07-743-357-6	Sequence 6, Appli
19	2779.5	61.0	875	4	US-09-952-060-6	Sequence 6, Appli
20	2774	60.9	850	4	US-09-952-060-2	Sequence 2, Appli
21	2755.5	60.5	875	4	US-09-952-060-8	Sequence 8, Appli
22	2750	60.4	850	4	US-09-952-060-4	Sequence 4, Appli
23	2746	60.3	562	3	US-09-117-217-14	Sequence 14, Appl
24	2746	60.3	562	4	US-09-735-487-14	Sequence 14, Appl
25	2715	59.6	560	4	US-10-205-641-1	Sequence 1, Appli
26	2464.5	54.1	1018	4	US-09-206-551-26	Sequence 46, Appl
27	2445.5	53.7	1016	4	US-09-625-972-24	Sequence 24, Appl
28	2216	48.6	441	4	US-09-991-258-16	Sequence 16, Appl
29	2173	47.7	1055	2	US-08-659-251-5	Sequence 5, Appli
30	2173	47.7	1055	3	US-09-256-490-5	Sequence 5, Appli
31	2173	47.7	1055	5	PCT-US96-11445-5	Sequence 5, Appli
32	2154.5	47.3	3080	6	5223423-4	Patent No. 5223423
33	2154.5	47.3	3080	6	5223423-4	Patent No. 5223423
34	2111	46.3	427	3	US-09-690-265-1	Sequence 1, Appli
35	1529	33.6	327	3	US-08-679-493A-72	Sequence 72, Appl
36	1512	33.2	327	3	US-08-679-493A-71	Sequence 71, Appl
37	1509	33.1	314	1	US-08-589-446-6	Sequence 6, Appli
38	1509	33.1	314	1	US-08-444-882-6	Sequence 6, Appli
39	1509	33.1	314	2	US-08-389-459A-6	Sequence 6, Appli
40	1509	33.1	314	3	US-08-987-867A-6	Sequence 6, Appli
41	1309	28.7	275	6	5252477-3	Patent No. 5252477
42	1309	28.7	275	6	5252477-3	Patent No. 5252477
43	1272	27.9	1150	3	US-09-238-303-9	Sequence 9, Appli
44	1272	27.9	1150	4	US-09-946-239-9	Sequence 9, Appli
45	1254	27.5	261	6	5320958-3	Patent No. 5320958

#### ALIGNMENTS

#### RESULT 1

US-07-743-357-1

; Sequence 1, Application US/07743357

; Patent No. 5858646

; GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; activity for use as diagnostic reagent and/or vaccine

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: KIRBY EADES GALE BAKER

; STREET: Box 3432, Station D

; CITY: Ottawa

; STATE: Ontario

; COUNTRY: Canada

; ZIP: K1M 1H8

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/07743,357

; APPLICATION NUMBER: US/07743,357

; FILING DATE: 21-AUG-1991

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA90/00062

; FILING DATE: 23-FEB-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Gale, Edwin J.

; REGISTRATION NUMBER: 28,584

; REFERENCE/DOCKET NUMBER: 30924-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (613) 237-6900

; TELEFAX: (613) 237-0045

; INFORMATION FOR SEQ ID NO: 1:

```
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1005 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;     ORGANISM: Human immunodeficiency virus type 1
;     STRAIN: HXB2
;   US-07-743-357-1

Alignment Scores:
Pred. No.:      9,89e-208      Length:      1005
Score:          3454.00      Matches:      651
Percent Similarity: 95.01%      Conservative: 34
Best Local Similarity: 90.29%      Mismatches: 28
Query Match:      75.83%      Indels:      8
DB:              2          Gaps:      4

US-09-610-313b-32 (1-2457) x US-07-743-357-1 (1-1005)

QY 220 TTCTTCCGCGAGACTGCGCTTCCCGAGGCAAGGCGCGAGTTCCTCCAGCGAGCAG 279
Db 2 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 21
QY 280 AACCGCGCAACAGCCCAACAGCCGAGCGAGTGCAGGTG-----CGCGGCGACAAACCCC 333
Db 22 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 41
QY 334 CGCAGCGAGGCGCGCGCGAGCGCGCGAGCGCGCGCGT-----AAGTTCCTCCAGATCACC 387
Db 42 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 61
QY 388 CTGTGCGCAGCGCCCTGTGTAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCGCTGCTG 447
Db 62 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 81
QY 448 GACACCGCGCGCGACACCGTGTGAGGAGATGAGCGTCCCGGCAAGTGGAGGCC 507
Db 82 AspThrGlyAlaAspThrValLeuGluMetSerLeuProGlyArgTrpLysPro 101
QY 508 AGATGATCGCGCGCATCGCGCTTCATCAGGTGGCGCGAGTACGACGACAGATCTGTATC 567
Db 102 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle 121
QY 568 GAGATCTCGGCAAGAAGCCATCGGCACCGCTGCTGATCGGCGCGCGCGCGTGAACATC 627
Db 122 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 141
QY 628 ATCGCGCGCAACATGTCAGCCAGTGGGTGTCACCTGAACTTCCCATCAGCGCCATC 687
Db 142 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 161
QY 688 GAGACGTCGCGTGAAGTGAAGCCCGCATGGAGCGCGCGCGCGCGCGTGAAGCGCGCC 747
Db 162 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 181
QY 748 CTGACCGAGGAGAAAGATCAAGCGCTGACCGCATCTCGAGGAGATGGAGAAGGAGGCG 807
Db 182 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 201
QY 808 AAGATCAACAAGATCGGCGCGCGAGAACCCCTTACAAACACCCCGTGTTCGCCATCAAGAAG 867
Db 202 LysIleSerLysIleGlyProGluAsnProTrpAsnThrProValPheAlaIleLysLys 221
QY 868 AAGGACACCAACAGTGGCGAGTGGTGGACTTCCGCGAGCTGAAACAGCGCGCACCCAG 927
Db 222 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 241
QY 928 GACTTCTGGAGGTGAGTGGGCGATCCCGCACCCCGCGCGCTCAAGAAAGAAAGAGAGC 987
Db
```

```
Db 242 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 261
QY 988 GTGACCGTGTGGAGCTGGGCGAGCGCTACTTACGCGTGGCCCTCCGAGGAGACTTCGCGC 1047
Db 262 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 281
QY 1048 AAGTACACCGCTTACCATCCCGAGCATCAACACGAGACCCCGCGCATCGCTACCAG 1107
Db 282 LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 301
QY 1108 TACAACGTGTGCGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 302 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 321
QY 1168 AAGATCTTGAGCCCTTCCGCGCGCGCGCGCGCGAGATCGTGATCTACCAG-----GCC 1221
Db 322 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 341
QY 1222 CCCCTGTACGTGGGCGAGCGACTCGAGATCGCGCAGCACCGCGCAAGATCGAGGAGCTG 1281
Db 342 AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 361
QY 1282 CGCAAGCACCTGTGCGCTGGGCGTTCACCAACCCCGCAAGAACACCAAGAGAGGCC 1341
Db 362 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 381
QY 1342 CCTTCTCTG-----CCCATCGAGCTGCACCCCGCACAGTGCAGCGTGCACCCCATCAG 1395
Db 382 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 401
QY 1396 CTGCGCGAGAGAGAGAGCTGGACCGTGAACACACATCCAGAAAGCTGGTGGCAAGCTGAAC 1455
Db 402 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 421
QY 1456 TGGGCGCAGCAGATCTACCCCGCGCATCAAGTGGCGCGAGCTGTGCAAGCTCTCGCGCGC 1515
Db 422 TrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuCysLysLeuLeuArgGly 441
QY 1516 GCGAAGGCGCTGACCGCATCGTCCCTGACCGAGGCGCGAGTGGAGCTGGAGCTGGCCGAG 1575
Db 442 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 461
QY 1576 AACCGCGAGATCTCGCGAGCGCGTGCACCGCTGTACTACGACCCCGCAGCAAGACCTG 1635
Db 462 AsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAspProSerLysAspLeu 481
QY 1636 GTGCGCGAGATCCAGAAGCGCGCGCACACAGTGGACCTACCATGAGATCTACCGAGAGGCC 1695
Db 482 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 501
QY 1696 TTCAGAAGACCTGAAGACCGGCAAGTACGCCAAGATGGCAGCGCGCGCGCACCAACGACGCTG 1755
Db 502 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 521
QY 1756 AAGCAGCTGACCGAGGCGGTGCAGAAGATCCCATGGAGAGCATCGTGATCTGGGCGCAAG 1815
Db 522 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 541
QY 1816 ACCCGCGAAGTTCGCGCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875
Db 542 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTrp 561
QY 1876 TGGCAGGCGCACTGATCCCGAGTGGAGTTCGTGAACACACCCCGCGCTGGTGAAGCTG 1935
Db 562 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 581
QY 1936 TGGTACCACTGAGAGAGAGCGCCATCATCGCGCGCGAGACTTCTACGTGAGCGCGCGCC 1995
Db 582 TrpTrpGlnLeuGluLysGluProIleValGlyAlaGluThrPheTrpValAspGlyAla 601
QY 1996 GCCAACCGCAGAGACCAAGATCGCAAGCGCGCTACGTGACCGAGCCCGGCGCGCGAGAG 2055
Db 602 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 621
```



Db 301 AenValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
QY 1171 ATCTTGGAGCCCTTCGGCCCGCAACCCCGAGATCGTATACAG-----CCCCC 1224  
Db 321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1225 CTGTACGTGGGAGCGACCTGGAGATCGCGACAGCCCGCAAGATCGAGGAGCTGGC 1284  
Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360  
QY 1285 AAGCACTGCTGCTGGGTTCACCAACCCCGCAAGAACACCAAGAGAGCCCGCC 1344  
Db 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
QY 1345 TTCCTG-----CCCATCGAGTGCACCCCGCAAGTGCAGCTGCAGCCATCGAGTG 1398  
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
QY 1399 CCGGAGAAGGAGAGCTGGACCGTGAACACATCCAGAAAGCTGGTGGCAAGCTGAACCTGG 1458  
Db 401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
QY 1459 GCCAGCCAGATCTACCCCGCATCAAGTGGCCAGCTGTGCAAGCTGCTGCGCGGCC 1518  
Db 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440  
QY 1519 AAGGCCCTGACGACATGTCGCCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAAC 1578  
Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuLeuAlaGluAsn 460  
QY 1579 CCGAGATCTCTCGGAGCGCTGCACGCGGTACTACGACCCCGCAAGAGCACTGTGTG 1638  
Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
QY 1639 GCCGAGATCCAGAAGCAGCGCCACACGACGTGACCTACCAATCTACGAGAGCCCTTC 1698  
Db 481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500  
QY 1699 AGAACCTTGAGACCGGCAAGTACCGCAAGATCGCCAGCGCCGACCAACCAAGCACTGAG 1758  
Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
QY 1759 CAGCTGACCGAGCGCTGCAGAAAGATCGCCATGGAGAGCATCGTATCTGGGGCAAGACC 1818  
Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
QY 1819 CCCAGTTCGCTCCCATCCAGAGAGACCTGGAGACCTGTGGACCGACTACTGTG 1878  
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
QY 1879 CAGGCCACTGTATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGCTGTGAGCTGTG 1938  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580  
QY 1939 TACCACTGGAGAAGAGCCCATCATCGCGCGGAGACTCTACGTGGACGCGCGCCGCC 1998  
Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 1999 AACCGCAGACCAAGATCGCAAGCGCGCTACGTACCGACCGACCGCGCGCGCAGAGATC 2058  
Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
QY 2059 GTGAGCTTGACCGACCAACCAAGACCGAGCTGCAGGCCATCCAGCTGCGCCCTG 2118  
Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2119 CAGGACAGCGGAGCGAGTGAACATCGTGACCGACGACGAGTACGCTGGGATCATC 2178  
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
QY 2179 CAGGCCCGCCCGCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGAGCTGATC 2238  
Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

QY 2239 AAGAAGGAGAGGTGTACTGAGCTGGTGCCCGCCCAAGGGCATCGGCGGCAACGAG 2298  
Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyAsnGlu 700  
QY 2299 CAGATCCGACAAAGCTGGTGAGCAAGGGCATCGCAAGGTGTCTCTCGACGCGCATCGAT 2358  
Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 3  
US-08-463-210-9  
; Sequence 9, Application US/08463210  
; Patent No. 6001977  
; GENERAL INFORMATION:  
; APPLICANT: CHANG, Nancy T.  
; APPLICANT: GALLO, Robert C.  
; APPLICANT: WONG-STAAI, Flossie  
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morgan & Finnegan, L.L.P.  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10154-0053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463.210  
; FILING DATE: 05-JUN-1995  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 06/693,866  
; FILING DATE: 23-JAN-1985  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 06/659,339  
; FILING DATE: 10-OCT-1984  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Serunian, Leslie A.  
; REGISTRATION NUMBER: 35,353  
; REFERENCE/DOCKET NUMBER: 2026-4193US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1015 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HTLV-III  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..1015  
; OTHER INFORMATION: /note= "pol protein of HTLV-III"

US-08-463-210-9

Alignment Scores:  
Pred. No.: 2,34e-206 Length: 1015  
Score: 3432.00 Matches: 650  
Percent Similarity: 93.18% Conservative: 33  
Best Local Similarity: 88.68% Mismatches: 20  
Query Match: 75.35% Indels: 20  
DB: 3 Gaps: 5

US-09-610-313B-32 (1-2457) x US-08-463-210-9 (1-1015)



Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 4

US-09-124-900-3

; Sequence 3, Application US/09124900

; Patent No. 6268484

; GENERAL INFORMATION:

; APPLICANT: KATINGER, Hermann

; APPLICANT: BUCHACHER, Andrea

; APPLICANT: ERNST, Wolfgang

; APPLICANT: BALLAUN, Claudia

; APPLICANT: PURTSCHER, Martin

; APPLICANT: TRKOLA, Alexandra

; APPLICANT: PREDL, Renate

; APPLICANT: SCHWATZ, Christine

; APPLICANT: KLINA, Annelies

; APPLICANT: STEINDL, Franz

; APPLICANT: MUSTER, Thomas

; TITLE OF INVENTION: HIV-Vaccines

; FILE REFERENCE: 1939-112P

; CURRENT APPLICATION NUMBER: US/09/124,900

; CURRENT FILING DATE: 1998-07-30

; PRIOR APPLICATION NUMBER: PCT/EP95/01481

; PRIOR FILING DATE: 1995-04-19

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Patentin version 3.0

; SEQ ID NO 3

; LENGTH: 1015

; TYPE: PRT

; ORGANISM: Human immunodeficiency virus type 1

US-09-124-900-3

Alignment Scores:

Pred. No.:	2,346-206	Length:	1015
Score:	3432.00	Matches:	650
Percent Similarity:	93.18%	Conservative:	33
Best Local Similarity:	88.68%	Mismatches:	30
Query Match:	75.35%	Indels:	20
DB:	3	Gaps:	5

US-09-610-313B-32 (1-2457) x US-09-124-900-3 (1-1015)

QY	220	TTCTTCCGCGAGACTCGCTTCCCGAGGCGCGAGTTC-----	267
Db	1	PhePheArgLysLeuAlaPheLeuGlnGlyLysAlaArgLysPheSerSerGluGln	20
QY	268	-----CCGAGCGAGCAGAACCGCGCCCAACAGCCCAACCCAGC	303
Db	21	ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg	40
QY	304	CCGAGCTGCGAGTG-----CGCGCGAGCAACCCCGCGAGCGAGCGCGCGCGCGCGC	357
Db	41	ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg	60
QY	358	CAGGCGACCCCTG-----AATCTCCCGAGATCACCTGTGCGCGCGCCCTGTGTGAGC	411
Db	61	GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr	80
QY	412	ATCAAGGTGGCGCGCGCATCAAGAGGCGCTTGTGGACACCGCGCGCGCGAGCACCCGCTG	471
Db	81	IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal	100
QY	472	CTGGAGGAGATGAGCTGCGCGCGAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGCGC	531
Db	101	LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly	120
QY	532	TTTCATCAGGTGCGCGCGTACACAGATCCTGTGATCGAGATCTGCGCGCAAGAGCCATC	591
Db	121	PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle	140
QY	592	GGCACCGTGTGATCGCGCGCCCAACCCCGTGAAACATCATCGCGCGCAACATGCTGACCCAG	651

Db	141	GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln	160
QY	652	CTGGCTGCACCCCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAG	711
Db	161	IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys	180
QY	712	CCCGCATGGACGCGCCCAAGGTGAAGCAGTGGCGCCCTGACGAGGAGAAAGATCAAGGCC	771
Db	181	ProGlyMetAspGlyProLysValysGlnTrpProLeuThrGluGluLysIleLysAla	200
QY	772	CTGACCGCCATCTCGAGGAGATGAGAGAGAGGCAAGATCACCAAGATCGGCCCGAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGlu	220
QY	832	AACCCCTACAACACCCCGCTGTTCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAG	891
Db	221	AsnProTyraAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys	240
QY	892	CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGCC	951
Db	241	LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
QY	952	ATCCCCCACCCCGCGCTCAAGAAGAAGAGCGTACCGTCTCGAGCTGGCGCAG	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp	280
QY	1012	GCCTACTTCAGGTGCGCCCTGACGAGGACTTCCGCAAGTACACCGCTTCCCATCCCT	1071
Db	281	AlaTyraPheSerValProLeuAspLysAspPheArgLysTyraThrAlaPheThrIlePro	300
QY	1072	AGCATCAACAACGAGACCCCGCGCATCGCTACAGTACAACAGTCTGCTCCCGAGGCTGG	1131
Db	301	SerIleAsnAsnGluThrProGlyIleArgTyraGlnTyraAsnValLeuProGlnGlyTrp	320
QY	1132	AAGGCGACCCCGCATCTTCAGAGCAGATGACCAAGATCTGAGACCTTCCGCGCC	1191
Db	321	LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys	340
QY	1192	CGCAACCCCGAGATCGTGATCTACAG-----GCCCGCTGTAGTGGCGAGCACCTG	1245
Db	341	GlnAsnProAspIleValIleTyraGlnTyraMetAspAspLeuTyraValGlySerAspLeu	360
QY	1246	GAGATCGCGCGACCGCGCAAGATCGAGAGCTGCGCAAGACCTGCTGCTGGTGGCGC	1305
Db	361	GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly	380
QY	1306	TTACACACCCCGCAAGAAGCACAGAGAGCGCCCTTCTCTG-----CCCATCGAG	1359
Db	381	LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyraGlu	400
QY	1360	CTGACACCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCGCGAGAGGAGAGCTGGACC	1419
Db	401	LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr	420
QY	1420	GTGAACGACATCCAGAAGCTGTGGCAAGTGAACCTGGCGCAGCAGATCTTACCCCGCGC	1479
Db	421	ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyraProGly	440
QY	1480	ATCAAGGTGCGCGCGCTGTGCAAGCTGCTGCGCGCGCAAGCGCTTCCCGACATCGTG	1539
Db	441	IleLysValArgGlnLeuLysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle	460
QY	1540	CCCTGACCGGAGGCGGAGCTGGAGTGGCGCGGAGACCGCGAGATCTCGCGCGAGCC	1599
Db	461	ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLysLysGluPro	480
QY	1600	GTGACCGCGTGTACTACGACCCCGCAAGGACCTGTGTGGCGCGAGATCCAGAAAGCGGC	1659
Db	481	ValHisGlyValTyraTyraProSerLysAspLeuIleAlaGluIleGlnLysGlnGly	500
QY	1660	CACGACAGTGGACCTTACAGATCTTACAGAGCGCTTCAAGACCTTCAAGACCGCGCAG	1719
Db	501	GlnGlyGlnTrpThrTyraGlnIleTyraGlnGluProPheLysAsnLeuLysThrGlyLys	520



```
QY 1720 TACGCCAAGATGCGCAGCCGCCACACCAACGAGCTGAAGCAGTGCACGAGCGCGTGCAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1780 AAGATCGCCATGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyThrProLysPheLysLeuProIle 560
QY 1840 CAGAAGGAGACCTGGGAGACCTGTGTGGACCACTACTGGCAGGCCACCTGATCCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpThrTrpGlnThrGlnAlaThrTripleProGlu 580
QY 1900 TGGAGTTCTGTGAACACCCCGCTGTGAAGCTGTGTACCACTGGAGAGAGAGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTrpGlnLeuGluLysGluPro 600
QY 1960 ATCATCGCGCGGAGACCTTCTAGTGGACCGCGCGCCCAACCGCGAGACCAAGATCGC 2019
Db 601 IleValGlyAlaGluThrPheTrpValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCCCGCTACGTGACCGCGCGCGCGCGGAGATCGTGAGCTGACCGAGACCAACC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACCAAGAGCGAGCTGACGGCCATCCAGCTGCGCTGCGAGGACGCGGAGCGAGGTG 2139
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCGTGCAGCGACGACCTAGCCCTGGGCGATCATCCAGGCCCGCGAGAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGACGAGCTGTGTGAACACGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTG 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCGCGCCCAAGGGGATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGlnGlnValAspLysLeuValSer 720
QY 2320 AAGGGCATCCGCAAGGTGCTCTTCTGACCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 5
US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STRAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866
```

```
; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1015
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
; US-08-463-028-9

Alignment Scores:
Pred No.: 2,34e-206 Length: 1015
Score: 3432.00 Matches: 650
Percent Similarity: 93.18% Conservative: 33
Best Local Similarity: 88.68% Mismatches: 30
Query Match: 75.35% Indels: 20
DB: 4 Gaps: 5

US-09-610-313B-32 (1-2457) x US-08-463-028-9 (1-1015)
QY 220 TTCTTCGCGAGGACCTGGGCTTCCCCCAGGGCAAGGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 268 -----CCCCGAGGAGAACCGCGCCCAACAGCCCGCCGAGCGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGGCGACAAACCCCGCAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGGCACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGTGGCGGCCAGATCAAGAGGCCCTGTGTGACACCGCGCGCGAGCAGCACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTGCGCGCAACTGGAAGCCCAAGATGATCGCGCGCATCGCGGC 531
Db 101 LeuGlnGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyIle 120
QY 532 TTCATCAAGGTGGCCAGTACGACAGATCTGTATCGAGATCTCGCGCAAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGGCCCGCCACCCCGTGAACATCATCGCGCCCGCAACATCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGACCCCTGACTTCCCATCATCGACCCCGAGCCCGCGCGCGTGAAGTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
```

Qy	712	CCCCGCATGGACGCGCCCAAGGTCAAGCAGTGGCCCTCTGACCGGAGGAGAAGATCAAGGCC	771
Db	181	ProGlyMetAspGlyProLeysValLysGlnTrpProLeuThrGluGluLysIleLysAla	200
Qy	772	CTGACCGCCCATCTCGGAGGAGATCGAGNAGGAGGGCAAGATCACCAGATCGGCCCGCAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGlnGlyLysIleSerLysIleGlyProGlu	220
Qy	832	AACCCCTACAAACACCCCGTGTTCGCATCAAGAAGAGGACACACCAAGTGGCGCAAG	891
Db	221	AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys	240
Qy	892	CTGTTGACATTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGC	951
Db	241	LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCCACCCCGCGCGCTGAAGAAGAGNAGAGCGTGACCGTGTGGAGCTGGGCGAC	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTACTTCAGCTGCGCCCTGGACGAGACTTCCGCAAGTACACCGCTTCACCATCCCC	1071
Db	281	AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACAGACACCCCGGCATCCGTTACCAGTACAACTGCTGCCACAGGCTGG	1131
Db	301	SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp	320
Qy	1132	AAGGGCAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCC	1191
Db	321	LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys	340
Qy	1192	CGCAACCCCGAGATCGTGATCTACCAG-----GCCCCCTGTACGTGGCGACGACCTG	1245
Db	341	GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu	360
Qy	1246	GAGATCGGCAGACACCGCGCAAGATCGAGGAGTGGCGAAGCACCTGCTCGCTGGCGGC	1305
Db	361	GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly	380
Qy	1306	TTACACACCCCGCAGACGACACAGACACAGAGAGCGCCCTTCTCTG-----CCCATCGAG	1359
Db	381	LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu	400
Qy	1360	CTGCACCCCGCAAGTGGACCGTGCACGCCATCGAGCTGCGCGAGAGGAGAGCTGCACC	1419
Db	401	LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr	420
Qy	1420	GTGAACGACATCCAGAAAGCTGTGGGCAAGCTGAACCTGGCGCACCCAGATCTACCCGCGC	1479
Db	421	ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly	440
Qy	1480	ATCAAGTGGCCAGCTGTGCAAGCTGTGCGCGCGCCAGCGCCCTGACCGACATCGTG	1539
Db	441	IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle	460
Qy	1540	CCCCTGACCGAGGAGCGGAGCTGGAGCTGCGCCGAGAACCGCGAGATCTCTGCGGAGCCC	1599
Db	461	ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluPro	480
Qy	1600	GTGCACGCGGTGTACTACGACCCAGCAAGACCTGTGTGGCGCGAGATCCAGAACGAGGC	1659
Db	481	ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly	500
Qy	1660	CACGACCAAGTGAACCTACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAG	1719
Db	501	GlnGlyGlnTrpTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys	520
Qy	1720	TACGCCAAGATCGCACCGGCCACCAACGACGTGAAGAGCTGACGAGGCGCGTCCGAG	1779
Db	521	TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln	540
Qy	1780	AAGATCGCCATGGAGAGCATCGTGCATCTGGGGCGCAAGACCCCAAGTTCGCGCTCCGCATC	1839

Db	541	LyseileThrThrGluSerileValIleTrpGlyLeuThrProLysPheLysLeuProIle	560
QY	1840	CAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCGCACTGGATGCCGAG	1899
Db	561	GlnLysGluThrTrpGluThrTrpThrGluTrpGlnAlaThrTrpIleProGlu	580
QY	1900	TGGGAGTTCTGTGAACACCCCGCCCTGGTGAAGCTGTGGTACCAGCTGGAGAGAGAGCCC	1959
Db	581	TrpGluPheValAsnThrProProLeuValLysLeuTrpGlnLeuGluLysGluPro	600
QY	1960	ATCATCGCGCCGAGACCTTCTAGTGGAGCGCGCGCCGACCAACCGCAGAGACCAAGATCGGC	2019
Db	601	IleValGlyAlaGluThrPheTyValAspGlyAlaAlaAsnArgGluThrLysLeuGly	620
QY	2020	AAGCGCGGTACTGTGACCGACCGGCGCGGCGCAGAAAGATCGTGAGCTGTGACCGAGACCA	2079
Db	621	LysAlaGlyTyThrValThrAsnLysGlyArgGlnLysValProLeuThrAsnThrThr	640
QY	2080	AACCAAGAAGACCGAGCTGCGAGCGCATCTGAGCTGGCCCTGCAGGACACGCGCAGAGG	2139
Db	641	AsnGlnLysThrGluLeuGlnAlaIleTyThrLeuAlaLeuGlnAspSerGlyLeuGluVal	660
QY	2140	AACATGTCGACCGACAGCGAGTACGCGCTGGGCGATCATCCAGGCGCGCGCAGCAAGAGC	2199
Db	661	AsnIleValThrAspSerGlnTyThrAlaLeuGlyIleGlnAlaGlnProAspLysSer	680
QY	2200	GAGAGCGAGCTGCTGAACAGATCATCGAGCAGCTCATCAAGAAGGAGAGGTGTACCTG	2259
Db	681	GluSerGluLeuValAsnGlnIleIleGluGlnLeuLysGluLysValTyLeu	700
QY	2260	AGCTGGTGCGCCGCCACAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGAGC	2319
Db	701	AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer	720
QY	2320	AAGGCGATCGCAAGTGCTGTTCTCGAGCGCATCGAT	2358
Db	721	AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp	733
RESULT 6			
US-07-743-357-2			
; Sequence 2, Application US/07743357			
; Patent No. 5858646			
; GENERAL INFORMATION:			
; APPLICANT: Kang, Yong C.			
; TITLE OF INVENTION: Polypeptide having immunological			
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine			
; NUMBER OF SEQUENCES: 22			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: KIRBY EADES GALE BAKER			
; STREET: Box 3432, Station D			
; CITY: Ottawa			
; STATE: Ontario			
; COUNTRY: Canada			
; ZIP: K1M 1H8			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: Patent In Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/07/743,357			
; FILING DATE: 21-AUG-1991			
; CLASSIFICATION: 424			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: PCT/CA90/00062			
; FILING DATE: 23-FEB-1990			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Gale, Edwin J.			
; REGISTRATION NUMBER: 28,584			
; REFERENCE/DOCKET NUMBER: 30924-2			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (613) 237-6900			

```

; TELFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1016 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   FRAGMENT TYPE: internal
;   ORIGINAL SOURCE:
;     ORGANISM: Human immunodeficiency virus type 1
;     STRAIN: BH102
;   US-07-743-357-2

Alignment Scores:
Pred. No.:      2,34e-206      Length:      1016
Score:          3432.00        Matches:      650
Percent Similarity: 93.18%      Conservative: 33
Best Local Similarity: 88.68%    Mismatches: 30
Query Match:      75.35%        Indels:      20
DB:               2           Gaps:         5

US-09-610-313B-32 (1-2457) x US-07-743-357-2 (1-1016)

Qy 220 TTCTTCGCGAGACCTGGCCCTTCCCGAGGCAAGGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAsePheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy 268 -----CCGAGCGAGAGAACCGCGGCAACAGCCGCGCAACAGCCCGCAGC 303
Db 21 ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg 40
Qy 304 CGCGAGCTGCAGGTG-----CGCGGCGACAAACCCCGCAGCGAGGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
Qy 358 CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCTCGTGAGC 411
Db 61 GlnGlyThrValSerPheAenPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
Qy 412 ATCAAGTGGCGCGCGAGATCAAGAGGCCCTGTGACACACCGCGCGCGAGCACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
Qy 472 CTGAGGAGATGAGCTGCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
Qy 532 TTCAATCAAGTGGCGCGAGTACGACAGATCCTGATCGAGATCTCGCGCAAGAGCCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
Qy 592 GGCACCGTGTGATCGGCGCGCACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
Qy 652 CTGGGCTGCACTTCCCTCCCATCGACCCCATCGAGCCGCGCGCGCTGAGCTGAG 711
Db 161 IleGlyCysThrLeuAenPheProIleSerProIleGluThrValProValLysLeuLys 180
Qy 712 CCGCGCATGAGCGCGCGCAAGGTGAAGCAGTGGCGCCCTCGCGAGGAGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
Qy 772 CTGACCGCCATCTCGGAGGAGATCGAAGAGGCGCAAGATCACCAAGATCGGCGCGCAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
Qy 832 AACCCCTACACACCCCGCTTCGCGCATCAAGAGAGAGACAGACCAAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240

```

```

Qy 892 CTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAGGAGCTTCTGGAGGTGCAGCTGGC 951
Db 241 LeuValAspPheArgGluLeuAenLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
Qy 952 ATCCCGCCACCGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysLysLysLysLysLysLysLysLysLys 280
Qy 1012 GCCTACTTCAGCGTGGCCCTGGAGAGAGACTTCGCGCAAGTACACCGCTTCACCATCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
Qy 1072 AGCATCAACAACAGAGACCCCGCGCATCCGTTACAGTACAGTGTGCTGCCCGAGGTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
Qy 1132 AAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
Qy 1192 CGCAACCCCGAGATCGTGATCTACCGAG-----GCCCGCCCTGTACGTGGCGAGCGACCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
Qy 1246 GAGATCGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAAGCAGCTGCTGCTGGCGC 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
Qy 1306 TTCAACACCCCGCAGAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1359
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
Qy 1360 CTGCACCCCGCAGAGTGGAGCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGAGTGCACC 1419
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
Qy 1420 GTGAACGACATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCCAGCCAGATCTACCCCGC 1479
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
Qy 1480 ATCAAGTGGCGCGAGCTGTGCAAGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1539
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
Qy 1540 CCCTTGACCGAGAGCGCGAGCTGGAGCTGCCGAGAACCGCGAGATCTCTGCGCGAGGCC 1599
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
Qy 1600 GTGCACGCGCTGTACTACGACCCCGCAGAGCCTGCTGGTGGCGCGAGATCCAGAGAGCG 1659
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
Qy 1660 CAGCAGCGTGGACCTTACAGATCTACAGAGCGCTTCAAGAACCTTGAAGACCGCGCAG 1719
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
Qy 1720 TAGCGCAAGATGCGCAGCGCCACACCAAGCTGAAGCAGCTGACCGAGCGCGCTGCAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
Qy 1780 AAGATCGCCATGGAGAGCATCTGTATCTGGGGCAAGACCCCGCGCTTCGCGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyThrProLysPheLysLeuProIle 560
Qy 1840 CAGAAGAGACCTGGGAGACCTGGTGGACCGCATCTATCGGAGAGCGCCACTGTGATCCCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTrpThrGluTrpGlnAlaThrTrpIleProGlu 580
Qy 1900 TGGGAGTTCGTGAAACCCCGCGCTGGTGAAGCTGGTACCAAGCTGGAGAGAGAGAGAGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpThrGlnLeuGluLysGluPro 600
Qy 1960 ATCATCGCGCGCGAGACCTTCTTACGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2019

```

[illegible]

## RESULT 7

```

US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LCMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-309-572-17

```

Alignment Scores:		
Pred. No.:	3.6e-206	Length: 1003
Score:	3429.00	Matches: 648
Percent Similarity:	94.53%	Conservative: 34
Best Local Similarity:	89.88%	Mismatches: 31
Query Match:	75.28%	Indels: 8
DB:	4	Gaps: 4
US-09-610-313B-32	(1-2457)	x US-09-309-572-17 (1-1003)

Qy	220	TTCCTCCGAGACCTGCCTCCCGCAGGCCAAGCCCGCGAGTTCCCACGCGACG	279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyysalaaargGluPheSerSerGluGln	20
Qy	280	AACCGCGCCAACAGCCCCACCAGCCGCGAGTGTGAGTG-----CGCGCGCAACAACCC	333
Db	21	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgaspAsnAsnSer	40
Qy	334	CGCAGCGAGCCGGCGCCGAGCGCCAGCGCACCTGTG-----AACTTCCCGCAGATCAC	387
Db	41	LeuSerGluAlaGlyAlaaspArgGlnGlyThrValSerPheSerPheProGlnIleThr	60
Qy	388	CTGTGCGACGCGCCCTGTGTAGCATCAAGTGTGGCGCGCCAGATCAAGAGAGGCCCTCTG	447

Db	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu	80
Qy	448	GACACCGCGCCGACGACACACCGTCTGGAGGAGATGAGCCTGCCGCGCAAGTGAAGCC	507
Db	81	AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyAArgTrpLysPro	100
Qy	508	AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCCAGTACGACACAGATCCTGATC	567
Db	101	LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTrpAspGlnIleLeuIle	120
Qy	568	GAGATCTCGCGCAAGAGGCGCATCGGCACCGTCTGATCGCGCCGCCACCCCGGTGAACATC	627
Db	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
Qy	628	ATCGCGCGCAACATGCTGACCAAGCTGGGTGCACCTGAACCTTCCCATCAGCGCCCATC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
Qy	688	GAGACCGTCCCGTGAAGTGAAGCCCGGCATGAGCGGCCCCAGGTGAAGCAGTGGGCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
Qy	748	CTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAGGAGGCC	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
Qy	808	AAGATCACCAAGATCGCGCCCGAGAACCCCTACAACACCCCGGTTCGCCATCAAGAAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220
Qy	868	AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCACCCAG	927
Db	221	LysAspSerThrLysTyrAspLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
Qy	928	GACTTCTGGAGGTGCGAGCTGGGCGATCCCGCCACCCCGCCGCTGAGAGAGAGAAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSer	260
Qy	988	GTGACCTGTGAGCGTGGCGCGCGCTACTCTAGCGTGCCTCTGGACGAGGACTCCGC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg	280
Qy	1048	AAGTACACCGCTTCACCATCCCGAGCATCAACAACGAGACCCCGCGCATCCGGTACCAAG	1107
Db	281	LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln	300
Qy	1108	TACAACTGTCTGCCAGGCGCTGGAGGGCAGCGCCAGCATCTTCCAGACGACGATGACC	1167
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr	320
Qy	1168	AAGATCTGTGAGCCCTTCGCGCGCGCAACCCCGAGATCTGTACTACCAAG-----GCC	1221
Db	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	340
Qy	1222	CCCCTGTACCTGGCGACGACCTGGAGATCGGCAGCAGCCGCGCAAGATCGAGGAGCTG	1281
Db	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
Qy	1282	CGCAAGACCTGTGCGCTGGGCTTCACCAACCCCGACAGAGACACCGAGAGAGGCC	1341
Db	361	ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
Qy	1342	CCCTTCCTG-----CCCATCGAGCTGCACCCGACAAAGTGGACCGTGGAGCCCATCGAG	1395
Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	400
Qy	1396	CTGCGCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGCGCAAGCTGAAC	1455
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
Qy	1456	TGGCGACGACGATCTACCCCGGCATCAAGGTGGCGCCAGCTGTGCAAGCTGCTCGCGGC	1515
Db	421	TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly	440

```
QY 1516 GCCAAGGCCCTGACCGACATCGTCCCTGACCGAGAGCGAGCTGAGCTGCGCCGAG 1575
Db 441 ThrLysAlaLeuThrGluValProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
QY 1576 AACCGGAGATCCCGCGAGCCGTCACCGCGTGTACTACGACCCCGAGCAAGACCTG 1635
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1636 GTGGCCGAGATCCAGAACGAGCGGCACACACAGTGGACCTACACAGATCTACGAGGCC 1695
Db 481 IleAlaGluIleGlnLysGlnGlyGlnThrThrThrThrGlnIleTyrGlnGluPro 500
QY 1696 TTCAGAACCTGAGAACCGCGAAGTACGCCAAGATGCGCACCGGCCACCAACGACGTG 1755
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY 1756 AAGCAGCTGACCGAGCGCGTCGACGAGATGCCATGGAGACATCGTATCTGGGGCAG 1815
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTyrGlyLys 540
QY 1816 ACCCCCAAGTCCCGCTCCCATCCAGAACGAGACCTGGGAGACCTGGTGACCGACTAC 1875
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTyrTrpThrGluTyr 560
QY 1876 TGGCAGGCCACTGATCCCGAGTGGAGTTCGTGAACACCCCGCCGACCTTCTACGTGGAGCGGCC 1935
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1936 TGCTACCACTGCGAGAGAGCCCATCATCCGCGCGGACCTTCTACGTGGAGCGGCC 1995
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 1996 GCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGCGCGCAGAG 2055
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY 2056 ATCGTGACCTGACCGAGACCAACCAACAGAACCGAGTGCAGGCGCATCCAGCTGGCC 2115
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2116 CTCGAGACAGCGCAGCGAGGTGAACATCGTGACCGACGACGACGACGCGCGGATC 2175
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2176 ATCCAGGCCCGCAGCAAGAGCGAGCGAGCTGGTGAACACGATCATCGACGACGTG 2235
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2236 ATCAAGAGGAGAGGTGTACCTGAGCTGGTGGTCCCGCCCAACAGGGGCATCGCGGCAAC 2295
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2296 GAGCAGATCGACAGCTGGTGAGCAAGGGCATCGCAAGGTGCTGCTCGGAGCGGATC 2355
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2356 GAT 2358
Db 721 Asp 721
```

## RESULT 8

```
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718.096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
```

```
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-718-096-17
```

```
Alignment Scores:
Pred. No.: 3,66-206 Length: 1003
Score: 3429,00 Matches: 648
Percent Similarity: 94,59% Conservative: 34
Best Local Similarity: 89,88% Mismatches: 31
Query Match: 75,28% Indels: 8
DB: 4 Gaps: 4
```

US-09-610-313B-32 (1-2457) x US-09-718-096-17 (1-1003)

```
QY 220 TTCTTCGCGAGGACCTGGGCTTCCCGCAGGCAAGCGCCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCACACGCCGCCACCGCGGAGCTGCAGGTG-----CGCGGCGACACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSer 40
QY 334 CGCAGCGAGGCGCGCGCGCGAGCGCGCAGCGCACCCCTG-----AACTTCCCGCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCTCGTGAGCATCAAGTGGCGGCCGACATCAAGAGCGCCCTGTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCAGCACCGCTGCTGAGGAGATGAGCTGCCCGGCAAGTGGNAGCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluLeuMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCAGTACGACGACCATCTGTATC 567
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeu 120
QY 568 GAGATCTGCGGCAAGAGGCCATCGGCACCGCTGTGATCGCGCCCGCCCGCTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATCGTCGACCCAGCTGGTGGTGCACCTGAACTTCCCATCAGGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCTGCGCGTGAAGCTGAAGCCGCGATGACCGCCCAAGGTGAAGCAGTGGGCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCCATCTGCGAGGAGATGGAGAGAGGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGAGAACCCCTAGACACACCCCGCTGTCGCCCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGGAGGTGCAGCTGGGCGCATCCCCCGCGCGCTGAAAGAAAGAGAGAGC 987
```

```

Db      241  AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLysGlnLysSer 260
QY      988  GTGACGCTGCTGGAGCGTGGGCGACCGCTACTTCAGCGTGCCCTCGAGCAGGACTTCGCG 1047
Db      261  ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspLysAspPheArg 280
QY      1048  AAGTACACCGCTTCACCATCCCGAGCATCAACAACGAGACCCCGGATCCGCTACACAG 1107
Db      281  LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 300
QY      1108  TACAACGCTGCTCCCGAGCGCTGGAAGGCGACCGCCAGCATCTTCAGACGACGATGACC 1167
Db      301  TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320
QY      1168  AAGATCTCGAGCCCTTCGCGCCCGCGCAACCCCGAGATCGTGATCTACAG-----TGC 1221
Db      321  LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTrpGlnTrpMetAsp 340
QY      1222  CCCCTGTAGTGGGAGCGACCTGGAGATCGGCACCGCGCCGCAAGATCGAGGAGCTG 1281
Db      341  AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 360
QY      1282  CGCAAGCACCTGCTGCGCTGGGGCTTCACACCCCGCGCAAGAGCAGACCAAGAGGCGCC 1341
Db      361  ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY      1342  CCTCTCTCTG-----CCCATCGAGTCGACCCCGCAAGTGACCGTCGAGCCCATCGAG 1395
Db      381  ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY      1396  CTGCGCGAGAGGAGGAGCTGGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAAC 1455
Db      401  LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY      1456  TGGGCGCAGCAGATCTACCCCGGCATCAAGGTGCGCGCAGCTGTCAAGTGTGCGCGCG 1515
Db      421  TrpAlaSerGlnIleTrpAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY      1516  GCCAGGCGCTGACGACACATCTGCGCTGACCGAGGAGCGCGAGCTGAGCTGCGCGAG 1575
Db      441  ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY      1576  AACCGCGAGATCTGCGCGAGCGCGTGCACGCGGTGTACTACGACCCAGCAAGGACCTG 1635
Db      461  AsnArgGluLeuLeuLysGluProValHisGlyValTrpTyrAspProSerLysAspLeu 480
QY      1636  GTGGCGCAGATCCAGAAGCAGGGCCACGACAGTGACCTACAGATCTTACCAGGAGCGCC 1695
Db      481  IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTrpGlnIleTrpGlnGluPro 500
QY      1696  TTCAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCGACCGCCGCGACCAACGACGTG 1755
Db      501  PheLysAsnLeuLysThrGlyLysTrpAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY      1756  AAGAGCTCACGAGCGCGTGCAGAGATCGCCATCGGAGAGATCGTGATCTGGGCGCAAG 1815
Db      521  LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540
QY      1816  ACCCCCAAGTTCGCGCTGCCCATCCAGAAAGGAGACCTGGGAGACCTGTGGACCGACTAC 1875
Db      541  ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTrp 560
QY      1876  TGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGTGTGAAGCTG 1935
Db      561  TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY      1936  TGTGTACACCTGAGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCG 1995
Db      581  TrpTrpGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTrpValAspGlyAla 600
QY      1996  GCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAGCGGGCGCGGAGAG 2055

```

```

Db      601  AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY      2056  ATCTGTAGCTGACCGCGAGACCAACCAACGAGAGCCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db      621  ValValProLeuThrAspThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY      2116  CTGACGAGACACGCGGAGGTGAACATCTGTGACCGACGACGACCACTACGCCCTGGGCATC 2175
Db      641  LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTrpAlaLeuGlyIle 660
QY      2176  ATCCAGGCCCGCGCGAGCGAAGAGGAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTG 2235
Db      661  IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY      2236  ATCAGAGGAGAGAGGTGTACTGAGCTGGTGGTGGCGCCGCGCAAGGATCGGCGGCAAC 2295
Db      681  IleLysLysGluLysValTrpLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
QY      2296  GAGCAGATCCGACAGCTGTGAGCAAGGCGCATCGCAGAGGTGCTGTCTTCTGGACGCGCATC 2355
Db      701  GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY      2356  GAT 2358
Db      721  Asp 721

RESULT 9
US-07-743-357-5
; Sequence 5, Application US/07743357
; Patent No. 5858646
; GENERAL INFORMATION:
; APPLICANT: Kang, Yong C.
; TITLE OF INVENTION: Polypeptide having immunological
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: KIRBY EADES GALE BAKER
; STREET: Box 3432, Station D
; CITY: Ottawa
; STATE: Ontario
; COUNTRY: Canada
; ZIP: K1M 1H8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/743,357
; FILING DATE: 21-AUG-1991
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/CA90/00062
; FILING DATE: 23-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gale, Edwin J.
; REGISTRATION NUMBER: 28,584
; REFERENCE/DOCKET NUMBER: 30924-2
; TELEPHONE: (613) 237-6900
; TELEFAX: (613) 237-0045
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1016 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: BRU

```

US-07-743-357-5

**Alignment Scores:**

Pred. No.:	5,56e-206	Length:	1016
Score:	3426.00	Matches:	648
Percent Similarity:	93.3%	Conservative:	36
Best Local Similarity:	88.4%	Mismatches:	29
Query Match:	75.21%	Indels:	20
DB:	2	Gaps:	5

US-09-610-313B-32 (1-2457) x US-07-743-357-5 (1-1016)

Qy	220	TTCTTCGGCAGGACCTGGCCCTTCCCCAGGCGCAAGCCCGCGAGTTC-----	267
Db	1	PhePheArgGluAAspLeuAlaPheLeuGluGlyLysAlaArgGluPheSerSerGluGln	20
Qy	268	-----CCCAGCGAGCAGAACCGCGCCCAACGAGCCCCCACCAGC	303
Db	21	ThrArgAlaAenSerProThrIleSerSerGluGlnThrArgAlaAenSerProThrArg	40
Qy	304	CGCAGAGCTGCAGGTG-----CGCGCGCACAACCCCGCAGCGAGCGCGCGCCAGCGC	357
Db	41	ArgGluLeuGlnValTrpGlyArgAspAenAenSerLeuSerGluAlaGlyAlaAspArg	60
Qy	358	CAGGCGACCTCG-----AATCTCCCCAGATCACCTGTGTGCAGCGCCCCCTGTGTGAGC	411
Db	61	GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr	80
Qy	412	ATCAAGGTGGCGGCGCAGATCAAGGAGCCCTGCTGGACACCGGCGCGCAGACACCGTG	471
Db	81	IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal	100
Qy	472	CTGAGAGAGATGAGCCTCCCGGCAAGTGGAAAGCCAAAGATGATCGCGGGCATCGCGCGC	531
Db	101	LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly	120
Qy	532	TTCATCAAGGTGGCGCAGTACGACAGATCCTGTATCGAGATCTGCGGCAAGAAGCCATC	591
Db	121	PheIleLysValArgGlnTyAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle	140
Qy	592	GGCACCGTGTGATCGGCGCCACCCCGTGAACATCATCGCGCCGCAACATGCTGTGACCCAG	651
Db	141	GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln	160
Qy	652	CTGGCGTGCACCTGAACTTCCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAG	711
Db	161	IleGlyCysThrLeuAenPheProIleSerProIleGluThrValProValLysLeuLys	180
Qy	712	CCCGGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTGACCGAGGAGAAGATCAAGGCC	771
Db	181	ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluLulLysIleLysAla	200
Qy	772	CTGACCGCCCATCTCGGAGGATGGAGAGGAGGGCAAGATCACCAAGATCGGCCCCGAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu	220
Qy	832	AACCCCTACACACCCCGTGTTCGCCATCAAGAAGAAGGACAGCACCAAGTGGCGCAAG	891
Db	221	AsnProTyranThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys	240
Qy	892	CTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGAGGTGCAGCTGGGC	951
Db	241	LeuValAspPheArgGluLeuAenLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCCACCCCGCGCTGAAGAAGAAGAGCGGTGACCGTGTCTGACGTGGCGGCAC	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTACTTCACGTGCGCCTCGACGAGGACTTCGCGAAGTACACCGCCTTCACCATCCCC	1071
Db	281	AlaTyxPheSerValProLeuAspGluAspPheArgLysTyxThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACACGAGACCCCGGCATCCGCTACCAAGTACAACGTGCTGCCCGCAGGCTGG	1131

Db	301	Ser	Ile	Asn	Asn	Glu	Thr	Pro	Gly	Ile	Arg	Tyr	Gln	Tyr	Asn	Val	Leu	Pro	Gln	Gly	Trp	320
Qy	1132	AAG	G	G	G	A	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	1191
Db	321	Lys	Gly	Ser	Pro	Ala	Ile	Phe	Gln	Ser	Ser	Met	Thr	Lys	Ile	Leu	Glu	Pro	Phe	Arg	Lys	340
Qy	1192	CG	CA	AC	CC	CC	G	A	T	CG	T	AC	C	A	G	-----	GC	CC	CC	CT	G	1245
Db	341	Gln	Asn	Pro	Asp	Ile	Val	Ile	Tyr	Gln	Tyr	Met	Asp	Asp	Leu	Tyr	Val	Gly	Ser	Asp	Leu	360
Qy	1246	GAG	AT	CG	CG	C	A	G	C	CG	CG	C	A	A	G	AT	CG	G	A	G	AT	1305
Db	361	Glu	Ile	Gly	Gln	His	Arg	Thr	Lys	Ile	Glu	Glu	Leu	Arg	Gln	His	Leu	Leu	Arg	Trp	Gly	380
Qy	1306	TT	C	A	C	A	C	C	C	C	C	C	A	A	G	A	G	A	G	C	C	1359
Db	381	Leu	Thr	Thr	Pro	Asp	Lys	Lys	His	Gln	Lys	Glu	Pro	Pro	Phe	Leu	Trp	Met	Gly	Tyr	Glu	400
Qy	1360	CT	G	A	C	C	C	C	C	A	A	G	T	G	C	A	C	C	C	C	A	1419
Db	401	Leu	His	Pro	Asp	Lys	Trp	Thr	Val	Gln	Pro	Ile	Val	Leu	Pro	Glu	Lys	Asp	Ser	Trp	Thr	420
Qy	1420	GT	G	A	A	C	A	T	C	C	A	A	G	CT	G	T	G	G	C	A	A	1479
Db	421	Val	Asn	Asp	Ile	Gln	Lys	Leu	Val	Gly	Lys	Leu	Asn	Trp	Ala	Ser	Gln	Ile	Tyr	Pro	Gly	440
Qy	1480	AT	C	A	A	G	T	G	C	C	A	G	T	G	C	T	G	C	G	G	C	1539
Db	441	Ile	Lys	Val	Arg	Gln	Leu	Cys	Lys	Leu	Leu	Arg	Gly	Thr	Lys	Ala	Leu	Thr	Glu	Val	Ile	460
Qy	1540	CCC	CT	G	A	C	C	A	G	A	G	G	C	C	G	A	G	A	C	C	G	1599
Db	461	Pro	Leu	Thr	Glu	Glu	Ala	Glu	Leu	Leu	Ala	Glu	Asn	Arg	Lys	Ile	Leu	Lys	Glu	Pro	480	
Qy	1600	GT	G	A	C	G	G	CT	GT	A	T	A	C	A	G	A	C	C	C	A	G	1659
Db	481	Val	His	Gly	Val	Tyr	Tyr	Asp	Pro	Ser	Lys	Asp	Leu	Ile	Ala	Glu	Ile	Gln	Lys	Gln	Gly	500
Qy	1660	C	A	G	A	C	A	G	T	G	A	C	T	A	C	C	A	G	G	C	C	1719
Db	501	Gln	Gly	Gln	Trp	Thr	Tyr	Gln	Ile	Tyr	Gln	Glu	Pro	Phe	Lys	Asn	Leu	Lys	Thr	Gly	Lys	520
Qy	1720	T	A	G	C	C	A	A	G	A	T	G	G	C	C	C	C	C	A	C	A	1779
Db	521	Tyr	Ala	Arg	Thr	Arg	Gly	Ala	His	Thr	Asn	Asp	Val	Lys	Gln	Leu	Thr	Glu	Ala	Val	Gln	540
Qy	1780	A	A	G	A	T	C	C	C	A	T	C	G	T	G	G	C	A	A	C	C	1839
Db	541	Lys	Ile	Thr	Thr	Glu	Ser	Ile	Val	Ile	Trp	Gly	Lys	Thr	Pro	Lys	Phe	Lys	Leu	Pro	Ile	560
Qy	1840	C	A	A	G	A	G	A	C	T	G	G	T	G	C	A	C	T	A	C	T	1899
Db	561	Gln	Lys	Glu	Thr	Trp	Glu	Thr	Trp	Trp	Thr	Glu	Tyr	Trp	Gln	Ala	Thr	Trp	Ile	Pro	Glu	580
Qy	1900	T	G	G	A	G	T	T	C	G	A	C	C	C	C	C	C	C	C	C	C	1959
Db	581	Trp	Glu	Phe	Val	Asn	Thr	Pro	Pro	Leu	Val	Lys	Leu									





Db 361 GluIleGlyGlnHisArgThrLysIleGluGluArgGlnHisLeuLeuArgTrpGly 380  
Qy 1306 TTCACCCCGCCGACGAGCACCAGAGAGAGCCCTTCCTG-----CCCATCGAG 1359  
Db 381 PheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTrpGlu 400  
Qy 1360 CTGACCCCGCCGACGAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGTGGACC 1419  
Db 401 LeuHisProAspLysTrpThrIleGlnProIleValLeuProGluLysAspSerTrpThr 420  
Qy 1420 GTGACGACATCCAGAGCTGGTGGGCAAGCTGAACCTGGCCGAGCAGATACCCCGGC 1479  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyProGly 440  
Qy 1480 ATCAAGGTGCCCGACGCTGTCAAGCTGTGGCGCGCGCAAGGCGCTGACCGACATCGTG 1539  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
Qy 1540 CCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCGCGAGGCC 1599  
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
Qy 1600 GTGACGCGCTGTACTAGACCCCGACGAGACCTGTGGCGGAGATCCAGAACGAGGC 1659  
Db 481 ValHisGlyValTyTrpAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
Qy 1660 CACGACGAGTGGACCTACGACATCTACGAGGAGCCCTTCAAGAACCTGAACACGCGAAG 1719  
Db 501 GlnGlyGlnTrpThrTyrglnIleTyrglnGluProPheLysAsnLeuLysThrGlyLys 520  
Qy 1720 TACGCCAAGATGCGCAGCCCGCCACACCAACGACGTGAAGCAGCTGACCGAGCCGTGCAG 1779  
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
Qy 1780 AAGATCGCATGGAGAGATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATC 1839  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
Qy 1840 CAGAGGAGACCTGGAGACCTGTGGACGACTACTGGCAGGCGACCTGGATCCCGCAG 1899  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluThrTrpGlnAlaThrTrpIleProGlu 580  
Qy 1900 TGGAGTTCGTGAACACCCCGCCCTGTGTAAGCTGTGTACAGCTGGAGAGAGAGCC 1959  
Db 581 TrpGluPheValAsnThrProLeuValLysLeuTrpTyrglnLeuLysGluPro 600  
Qy 1960 ATCATCGCGCGGAGACCTTCTAGTGAACGCGCGCGCCCAACCGCGAGACCAAGATCGG 2019  
Db 601 IleValGlyAlaGluThrPheTyrrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620  
Qy 2020 AAGCGCGCTACGTGACCGCGCGCGCGCAGAGATCGTGAGCCTGACCGACCCACC 2079  
Db 621 LysAlaGlyTyrrValThrAsnArgGlyArgGlnLysValValThrLeuThrHisThr 640  
Qy 2080 AACGAGAGACCGAGCTGCAGCGCATCTGAGTGGCCCTGACGAGCAGCGCGAGCGAGTG 2139  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
Qy 2140 AACATCGTGACCGACGACGATGACGCTGGGCAATCATCGAGGCCCGCCGACGAGC 2199  
Db 661 AsnIleValThrAspSerGlnTyrrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
Qy 2200 GAGAGCGAGCTGTGACCGACATCATCGACGAGCTGATCAAGAGGAGAGGTTACTGTG 2259  
Db 681 GluSerGluLeuValAsnGlnIleIleGlnLeuIleLysLysGluLysValTyrrLeu 700  
Qy 2260 AGCTGGGTGCGCGCCCAACAGGGCATCGCGCGCAACGAGCAGATCGCAAGCTGCTGAGC 2319  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720  
Qy 2320 AAGGCGATCCGCAAGGTGCTGTCTCGACGCGCATCGAT 2358  
|||||

Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733  
RESULT 11  
US-07-743-357-4  
; Sequence 4, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1016 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: PV22  
; US-07-743-357-4  
Alignment Scores:  
Pred. No.: 9,87e-206 Length: 1016  
Score: 3422.00 Matches: 647  
Percent Similarity: 93.18% Conservative: 36  
Best Local Similarity: 88.27% Mismatches: 30  
Query Match: 75.13% Indels: 20  
DB: 2 Gaps: 5  
US-09-610-313B-32 (1-2457) x US-07-743-357-4 (1-1016)  
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTC----- 267  
Db 1 PhePheArgLysAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
Qy 268 -----CCGACGAGCAGAACCCGCGCAACAGCCACAGCCACCGCAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
Qy 304 CGCGAGCTGAGGTG-----CGCGCGCACACCCCGCAGCGAGCGCGCGCGCGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
|||||



ADDRESSEE: KIRBY EADES GALE BAKER  
 STREET: Box 3432, Station D  
 CITY: Ottawa  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: K1M 1H8  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/743,357  
 FILING DATE: 21-AUG-1991  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA90/00062  
 FILING DATE: 23-FEB-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gale, Edwin J.  
 REGISTRATION NUMBER: 28,584  
 REFERENCE/DOCKET NUMBER: 30924-2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (613) 237-6900  
 TELEFAX: (613) 237-0045  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1003 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 FRAGMENT TYPE: internal  
 ORIGINAL SOURCE:  
 ORGANISM: Human immunodeficiency virus type 1  
 STRAIN: ELI  
 US-07-743-357-10

## Alignment Scores:

Pred. No.: 2,17e-205 Length: 1003  
 Score: 3416.50 Matches: 642  
 Percent Similarity: 94.72% Conservative: 40  
 Best Local Similarity: 89.17% Mismatches: 31  
 Query Match: 75.01% Indels: 7  
 DB: 2 Gaps: 4

US-09-610-313b-32 (1-2457) x US-07-743-357-10 (1-1003)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279  
 DB 1 PhePheArgGluAenLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerProLysGln 20  
 QY 280 AACCGCCACACGCCCCACCGAGCGGAGTGGCGCGC-----GACACCCCGCG 336  
 DB 21 ThrArgAlaAenSerProThrSerArgGluLeuArgValTrpGlyArgAspAenProLeu 40  
 QY 337 AGCGAGGCGCGCGCGAGCGCGAGGCGACCCCTG-----AACTTCCCGAGATCACCCCTG 390  
 DB 41 SerLysThrGlyAlaGluArgGlnGlyThrValSerPheAenPheProGlnIleThrLeu 60  
 QY 391 TGGCAGCGCCCTGGTGAGCATCAAGTGGCGGCGCAGATCAAGAGGCGCCCTGCTGGAC 450  
 DB 61 TrpGlnArgProLeuValAlaIleLysIleGlyGlnLeuLysGluAlaLeuAsp 80  
 QY 451 ACCGCGCCGACGACCGGTGCTGGAGAGATGAGCTGCCCGCAAGTGGAGCCCAAG 510  
 DB 81 ThrGlyAlaAspAspThrValLeuGluGluMetAenLeuProGlyLysTrpLysProLys 100  
 QY 511 ATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGCAGTACGACACGATCCTGATCGAG 570  
 DB 101 MetIleGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProIleGlu 120

QY 571 ATCTGCGGCAAGAGGCGCATCGGCACCGTGTGATCGGCCCCACCCCGTGAACATCATC 630  
 DB 121 IleCySGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAenIleIle 140  
 QY 631 GCGCGCAACATGCTGACCCAGCTGGCGGTGGACCTGAACCTCCCATCAGCCCATCAG 690  
 DB 141 GlyArgAenLeuThrGlnIleGlyCysThrLeuAenPheProIleSerProIleGlu 160  
 QY 691 ACCGTGCCGTGAAGCTGAAGCCCGCATCGACCGGCCCAAGGTGAAGACAGTGGCCCTG 750  
 DB 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
 QY 751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCCATCTCGGAGGAGATGAGAAAGAGGCGAAG 810  
 DB 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys 200  
 QY 811 ATCACCAGATCGGCCCGGAGAACCCCTACAAACCCCGTGTGTCGCCATCAAGAAGAG 870  
 DB 201 IleSerArgIleGlyProGluAenProIlePheAenIleLysLysLys 220  
 QY 871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGAGCTGAACAAAGCGCACCCAGGAC 930  
 DB 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAenLysArgThrGlnAsp 240  
 QY 931 TTCTGGAGGTGACGTGGGCATCCCCACCCCGCGCTGAAGAAAGAGAGCGGTG 990  
 DB 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
 QY 991 ACCGTGCTGACGTGGCGGACGCTACTTCAAGCGTGGCCCTGGAGGAGTCTCCGCAAG 1050  
 DB 261 ThrValLeuAspValGlyAspAlaIlePheSerValProLeuAspGluAspPheArgLys 280  
 QY 1051 TACACCGCCTTCCACCATCCCGCAGCATCAACAAAGAGACCCCGCATCCGCTACAGTAC 1110  
 DB 281 TyrThrAlaPheThrIleSerSerIleAenAenGluThrProGlyIleArgTyrGlnTyr 300  
 QY 1111 AAGTGTGCTGCCAGGCTGGAAGGCGAGCGCCCGAGCATCTTCCAGAGCAGCATGACCAAG 1170  
 DB 301 AenValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTGGAGCGCTTCGCGCGCCCGCAGACCCCGAGATCGTGATCTACAG-----GCCCCC 1224  
 DB 321 IleLeuGluProPheArgLysGlnAenProGluMetValIleTyrGlnTyrMetAspAsp 340  
 QY 1225 CTGTACGTGGCGCAGCACCTGGAGATCGGCAGCACCGCGCGCAAGATCGAGGAGCTGCGC 1284  
 DB 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg 360  
 QY 1285 AAGCAGCTGTGCTGGCGGCTTCCACACCCCGCAAGAGCAAGAGAGAGGAGCCCGCC 1344  
 DB 361 GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro 380  
 QY 1345 TTCTGT-----CCCATCGAGCTGACCCCGAGAGTGGACCGTGGAGCCCATCGAGCTG 1398  
 DB 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400  
 QY 1399 CCGAGAGAGAGAGTGGACCGGTGAACACATCCAGAAAGCTGGTGGCGCAAGCTCAACTGG 1458  
 DB 401 ProGluLysGluSerTrpThrValAenAspIleGlnAenLeuValGluArgLeuAenTrp 420  
 QY 1459 GCCAGCCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCC 1518  
 DB 421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440  
 QY 1519 AAGCGCTGACGACATCGTGGCCCTGACCGAGGAGCGCGAGTGGAGCTGGCGGAGAAC 1578  
 DB 441 LysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGluAen 460  
 QY 1579 CGCGAGATCTCGCGAGCGCGTGCACGGGTGTACTACGACCCCGACAGAGGAGCTGTGTG 1638  
 DB 461 ArgGluLysLeuLysGluProValHisGlyValTyrTrpAspProSerLysAspLeuIle 480  
 QY 1639 GCCGAGATCCAGAGAGCGGCCACGACCGAGTGGACCTTACAGATCTTACAGGAGCCCTTC 1698

Db 481 AlaGluLeuGlnGlyGlnGlyHisGlyGlnTrpThrGlnLeuProPhe 500  
QY 1699 AAGAACCTGAAGACCGGCAAGTACCCCAAGATGCGCCACCAACCAAGCTGAAG 1758  
Db 501 LysAsnLeuLeuThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVallys 520  
QY 1759 CAGCTGACCGAGCGCTGAGAGATGCCATGCGAGAGATCGTGTATCTGGGCAAGACC 1818  
Db 521 GlnLeuAlaGluAlaValGlnArgLeuSerThrGluSerIleValIleTrpGlyArgThr 540  
QY 1819 CCCAAGTTCGCGCTCCCATCCAGAGGAGACCTGGGAGACCTGTGGACCGACTACTGG 1878  
Db 541 ProLysPheArgLeuProLleGlnLysGluThrTrpGluThrTrpAlaGluLysTrp 560  
QY 1879 CAGCGACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGG 1938  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
QY 1939 TACGAGCTGGAGAACGACCCATCATCGCGCGGAGACCTTCTAGCTGAGCGCGCCGCC 1998  
Db 581 TyrGlnLeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 1999 AACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGCGCGCGCGGAGATC 2058  
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
QY 2059 GTGAGCTGACCGAGACCAACCAAGACCGAGTGGAGCGGATCCAGCTGGCGCTG 2118  
Db 621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640  
QY 2119 CAGGACGCGGAGAGGTGACATCGTACCGGACCGGAGTACCGCTGGCGATCATC 2178  
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
QY 2179 CAGGCGCAGCGCGCAAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGAGCTGATC 2238  
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleGluLeuLeu 680  
QY 2239 AAGAGGAGAGGTACTGAGCTGGTGGTGGCGCGCGCGGATCGGCGGCAACGAG 2298  
Db 681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
QY 2299 CAGATCGACAGCTGGTGAAGAGGCGATCCGAGAGTGTCTTCTGAGCGGATCGAT 2358  
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

## RESULT 13

US-07-743-357-8  
; Sequence 8, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY BADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: KIM 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1003 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: RF  
; US-07-743-357-8

Alignment Scores:  
Pred. No.: 3,34e-205 Length: 1003  
Score: 3413.50 Matches: 643  
Percent Similarity: 94.72% Conservative: 39  
Best Local Similarity: 89.31% Mismatches: 31  
Query Match: 74.94% Indels: 7  
DB: 2 Gaps: 4

US-09-610-313B-32 (1-2457) x US-07-743-357-8 (1-1003)

QY 220 TTCTTCCGCGAGACCTGGCTTCCCCAGGCAAGGCCCGAGTTCCTCCAGGAGCAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluLeuSerSerGluGln 20  
QY 280 AACCGCGCAACACGCCCCCAGCGCGGAGTGTGAGTGGCGGCGC--GACACCCCGC 336  
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40  
QY 337 AGCAGGCGCGCGCGAGCGCGAGCGGACCGCTG-----AACTTCCCCCAGATCACCTG 390  
Db 41 SerGluAlaGlyGluAspArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60  
QY 391 TGGCAGCGCCCCCTGGTGTGAGCATCAAGTGGCGCGCGCAGATCAAGGAGGCGCTGTGAC 450  
Db 61 TrpGlnArgProIleValThrValLysIleGlyGlnLeuLysGluAlaLeuAsp 80  
QY 451 ACCGCGCGCGAGACCGCTGTGAGGAGATGAGCTGCGCGGCGAGTGGAGGCCCAAG 510  
Db 81 ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100  
QY 511 ATGATCGCGCGCATCGCGGCTTTCATCAAGTGGCGCGCAGTACGACCATCTGTATCGAG 570  
Db 101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120  
QY 571 ATCTCGGCAAGAGGCGCATCGGACCGTGTGTGTGTCGCGCGCGCGCGCGTGAACATCATC 630  
Db 121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140  
QY 631 GGCGGCAACATGTCACCGCTGCGGCTGCGACCTGACTTCCCATCAGCCCATCGAG 690  
Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
QY 691 ACCGTGCGCGTGAAGCTGAAGCCCGGATGACCGCGCGCGCGCGCGCGCGCGCGCTG 750  
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
QY 751 ACCGAGGAGAGATCAAGCGCCCTGACCGCCCATCTTGGAGGAGATGGAGAGAGGCGCAAG 810  
Db 181 ThrGluGluLysIleLysAlaLeuValIleCysThrGluMetGluLysGluGlyLys 200

QY 811 ATCAACAGATCGGCGCCGAGAACCCCTTACCAACACCCCGTGTTCGCATCAAGAAAG 870  
DB 201 ILeSerLySileGlyProGluAAsnProTyrAsnThrProValPheAlaIleLySLeYs 220  
QY 871 GACAGCAACCAAGTCGCGAAGCTGCTGAGCTTCGCGAGCTGAACAGCCGACCCAGGAC 930  
DB 221 AspSerThrLySAspThrPheGlyLeuValAspPheArgGluLeuAAsnLySAspThrGlnAsp 240  
QY 931 TTCTGGAGGTGACAGCTGGGATCCCCACCCCGCCGCTTGAAGAAGAAAGAGCGTG 990  
DB 241 PheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLySLeYsLeYsSerVal 260  
QY 991 ACCGTGCTGAGCTGGGAGCGCTTACTTACGCTGCGCTGAGAGAGACTTCGCGAAG 1050  
DB 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLySLeYsLeYsLeYs 280  
QY 1051 TACACCGCTTACCATCCCGAGCATCAACCAACAGACCCCGCATCCGCTACCATAC 1110  
DB 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgIleArgTyrGlnTyr 300  
QY 1111 AACGTGCTGCCCGCAGGCTGGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATACCAAG 1170  
DB 301 AsnValLeuProGlnGlyTrpLySLeYsGlySerProAlaIlePheGlnSerSerMetThrLyS 320  
QY 1171 ATCTGAGGCTTCCGCGCCCGCAACCCCGAGATCGTATACAG-----GCCGCC 1224  
DB 321 IleLeuGluProPheLySLeYsGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1225 CTGTACCTGGCAGCGAGCTGGAGATCGGCGAGCAGCCGCGCAGATCGAGGACTCGC 1284  
DB 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLySLeYsLeYsLeYsLeYs 360  
QY 1285 AAGCACTGCTGCTGGGCTTCAACACCCCGCAGAGAACCAACAGAGGAGCCCGCC 1344  
DB 361 GluHisLeuLeuTyrTrpGlyPheThrThrProAspLySLeYsGlnLySLeYsLeYsLeYs 380  
QY 1345 TTCTCTG-----CCATCGAGCTCACCCCGCAGAGTGGACCGTGGAGCTGCAGCTG 1398  
DB 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLySLeYsLeYsLeYsLeYsLeYsLeYs 400  
QY 1399 CCGAGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGCTGGCGAGCTGAATCG 1458  
DB 401 ProGluLySAspSerTrpThrValAsnAspIleGlnLySLeYsLeYsLeYsLeYsLeYs 420  
QY 1459 GCCAGCAGATCTACCCCGGATCAAGTGCGCGAGCTGTCAAGCTGTGCGGGCGCC 1518  
DB 421 AlaSerGlnIleTyrAlaGlyIleValLySLeYsLeYsLeYsLeYsLeYsLeYsLeYs 440  
QY 1519 AAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGAGAAC 1578  
DB 441 LySAlaLeuThrGluValValGlnLeuThrLySLeYsGluAlaGluLeuAlaGluAsn 460  
QY 1579 GCGAGATCTGCGCGAGCGCGTGCAGCGGTGTACTACGACCCCGAGAGGAGCTGGTG 1638  
DB 461 ArgGluIleLeuLySLeYsGluProValHisGlyValTyrTyrAspProSerLySAspLeuIle 480  
QY 1639 GCCAGATCCAGAGAGGCGCACAGCAGTGGACCTACAGATCTACAGAGCGCCCTTC 1698  
DB 481 AlaGluIleGlnLySLeYsGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500  
QY 1699 AAGAACTGAAGACCGGCAAGTACGCAAGATGGCAGCCCGCCACACCAACAGAGTGAAG 1758  
DB 501 LySAsnLeuLySLeYsLeYsTyrAlaArgMetArgGlyAlaHisThrAsnAspValLyS 520  
QY 1759 CAGTGAACCGAGCGCTGCGAGAAGATGCCATGAGAGAGCATCGTGTGCTGGGCAAGACC 1818  
DB 521 GlnLeuThrGluAlaValGlnLySValAlaThrGluSerIleValIleTrpGlyLySThr 540  
QY 1819 CCCAAGTTCGCTGCCCTTCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878  
DB 541 ProLySLeYsLeuProIleGlnLySLeYsLeYsLeYsLeYsLeYsLeYsLeYsLeYs 560  
QY 1879 CAGGCCACCTGGATCCCGCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGG 1938

DB 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLySLeuTrp 580  
QY 1939 TACCAGCTGAGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC 1998  
DB 581 TyrGlnLeuGlnLySLeYsGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 1999 AACCGCAGAGCAAGATCGGCAAGCGCGGTGTACGTGACCGACCGGGCGCGCAGAAGATC 2058  
DB 601 AsnArgGluThrLySLeYsLeYsAlaGlyTyrValThrAspArgGlyArgGlnLySVal 620  
QY 2059 GTGAGCCTGACCGAGAGACCAACCAAGACCGAGCTGCGGCGCATCCAGCTGGCCTG 2118  
DB 621 ValSerLeuThrAspThrThrAsnGlnLySLeYsThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2119 CAGGACAGCGCGCAGCAGGTGAAACATCGTGACCGACAGCAGTAGTACGCCCTGGGCATCATC 2178  
DB 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
QY 2179 CAGCCCGCAGCCGCAAGAGCAGAGCGAGCTGTGAAACCAAGATCATCGAGCAGCTGATC 2238  
DB 661 GlnAlaGlnProAspLySLeYsSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680  
QY 2239 AAGAGAGAGAGGTGTACCTGAGCTGGTGGTGGCGCCCGCAGAGGCGCATCGCGGCAACGAG 2298  
DB 681 LySLeYsGlnLySValTyrLeuAlaTrpValProAlaHisLySLeYsLeYsLeYsLeYs 700  
QY 2299 CAGATCCACAGCTGTGTGAGCAGGCGCATCGCAAGGTGTGTCTTCTGGAGCGCATCGAT 2358  
DB 701 GlnValAspArgLeuValSerThrGlyIleArgLySValLeuPheLeuAspGlyIleAsp 720

## RESULT 14

US-07-743-357-7

; Sequence 7, Application US/07743357

; Patent No. 5858646

; GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: KIRBY EADES GALE BAKER

; STREET: Box 3432, Station D

; CITY: Ottawa

; STATE: Ontario

; COUNTRY: Canada

; ZIP: K1M 1H8

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07743,357

; FILING DATE: 21-AUG-1991

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA90/00062

; FILING DATE: 23-FEB-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Gale, Edwin J.

; REGISTRATION NUMBER: 28,584

; REFERENCE/DOCKET NUMBER: 30924-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (613) 237-6900

; TELEFAX: (613) 237-0045

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1004 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

```

;
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
;
; ORIGINAL SOURCE:
;
; ORGANISM: Human immunodeficiency virus type 1
; STRAIN: SF2
; US-07-743-357-7

```

Alignment Scores:		
Pred. No.:	9.82e-205	1004
Score:	3406.00	642
Percent Similarity:	94.31%	Conservative: 38
Best Local Similarity:	89.04%	Mismatches: 33
Query Match:	74.77%	Indels: 8
DB:	2	Gaps: 4

US-09-610-313B-32 (1-2457) x US-07-743-357-7 (1-1004)

Qy	220	TTCTTCGCGAGGACTGTGGCTTCCCCAGGCGAAGCCCGCGAGTGTCCCGACGCGAGCAG	279
Db	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
Qy	280	AACCGCGCCACAGCCCCACAGCCGCGAGCTGCAGTGC CGGC-----GACAACCCC	333
Db	21	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTprGlyGlyGluAAsnAsnSer	40
Qy	334	CGCAGCAGGCGCGCGCGCGAGCGCCAGGCGACCCCTG-----AACTTCCCGCCAGATCACC	387
Db	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr	60
Qy	388	CTGTGGCAGCGCCCTCTGTGTAGAGATCAAGTGGCGCGCCAGATCAAGGAGGCCCTGTCTG	447
Db	61	LeuTprGlnArgProLeuValThrIleArgIleGlyGlyLeuLeuLysGluAlaLeuLeu	80
Qy	448	GACACCGCGCGCAGACACCGTCTCGAGAGATGAGCTGCCGCGCAAGTGGAGAGCCC	507
Db	81	AspThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyIleTprLysPro	100
Qy	508	AAGATGATCGCGGCATCGCGGGCTTCATCAAGTGGCGCCAGTACGACCCAGATCTCTGATC	567
Db	101	LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal	120
Qy	568	GAGATCTCGCGCAAGAGGCCATCGGCACCGTCTGATCGCGCCCGCCACCCCGGTGACATC	627
Db	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
Qy	628	ATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCGCATCAGCCCATC	687
Db	141	IleGlyArgAsnLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
Qy	688	GAGACCGTGCCGTGAAGCTGAACCCCGGCATGAGCGGCCCGCCAAAGGTGAAGCAGTGGCCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTprPro	180
Qy	748	CTGACCGAGGAGAAGATCAAGGCCCTCACCGCCATCTCGAGAGATGGAGAGAGGAGGCG	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
Qy	808	AAGATCACCAAGATCGGCGCCCGAGACCCCTTACAACACCCCGGTGTTCGCCATCAAGAAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220
Qy	868	AAGACACACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAGCGCACCCAG	927
Db	221	LysAspSerThrLysTprArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
Qy	928	GACTTCTGGGAGTGTGAGCTGGGCATCCCCCACCCCGCGCTTGAAGAAGAAGAAGAGC	987
Db	241	AspPheTprGluValGlnLeuGlyIleProHisProalGlyLeuLysLysLysLysSer	260
Qy	988	GTGACCGTGCTGAGCTGGGCGACGCTTACTTTCAGCTGCGCCCTGTGACGAGGACTTCCGC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysLysPheArg	280



```
Db 641 LeuGlnAepSerGlyLeuGluValAsnIleValThrAepSerGlnTyrAlaLeuGlyIle 660
Qy 2176 ATCCAGGCCAGCCGACAGAGCCAGAGCGAGTGGTGAACAGATCATCGAGCAGCTG 2235
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy 2236 ATCAAGAAGGAGAGGTGTACTGCTGAGCTGGTGGCCGCCCCCAAGAGGATCGCGGCAAC 2295
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
Qy 2296 GAGCAGATCGACAAGCTGGTCAGCAAGGGCATCCGCAAGGTGCTGCTTCCTCGAGCGGCATC 2355
Db 701 GlnGlnValAepLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle 720
Qy 2356 GAT 2358
Db 721 Asp 721

RESULT 15
US-09-319-588C-6
; Sequence 6, Application US/09319588C
; Patent No. 6509018
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAUCLERE, Philippe
; APPLICANT: LOUSSERT-AJAXA, Ibtissam
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSI, Françoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/09/319,588C
; CURRENT FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-319-588C-6

Alignment Scores:
Pred. No.: 3,72e-188 Length: 1014
Score: 3140.50 Matches: 579
Percent Similarity: 89.70% Conservative: 74
Best Local Similarity: 79.53% Mismatches: 60
Query Match: 68.95% Indels: 15
DB: 4 Gaps: 5

US-09-610-313B-32 (1-2457) x US-09-319-588C-6 (1-1014)
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCCAGGCGCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProProAspAsn 20
Qy 280 AAC-----CGCGCAACAGCCCAAGCGCGCGCGAGTGCAGGTG----- 318
Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40
Qy 319 ---CGCGCGCAACCCCGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60
Qy 367 CTGAATCTCCCGCAGATCACCCTGTGGCAGCGCCCTGGTGAGCATCAAGGTGGCGGC 426
Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
Qy 427 CAGATCAAGGAGGCGCCCTGCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486
```

```
Db 81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluLeuGln 100
Qy 487 CTGCCCCGCAAGTGGAAAGCCCAAGATCATCGCGGCATCGCGGCTTCATCAAGGTGGCG 546
Db 101 LeuGluGlyLysTrpLysProLysMetIleGlyIleGlyLysPheIleLysValArg 120
Qy 547 CAGTACACCAAGATCTGTAGATCTGCGGCAAGAGGCATCGGCACCGTGTGTATC 606
Db 121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140
Qy 607 GGGCCCCACCCCGTGAACATCATCGCGGCACATGCTGACCCAGCTGGGTGGTGCACCTG 666
Db 141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160
Qy 667 AACTTCCCATCATCGCCCATCGAGACCGTCCCTGAAGCTGAAGCCCGGCATCGACGGC 726
Db 161 AsnPheProLysSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180
Qy 727 CCCAAGGTGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGC 786
Db 181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200
Qy 787 GAGGAGATGAGAAGGAGGAGGAGATCACCAGATCGGCCCGCGAGAACCCCTACACACC 846
Db 201 ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220
Qy 847 CCGTGTTCCTCCATCAAGAAGAGAGGAGCAGCACCAGTGGCGCAAGCTGGTGGACTCCGC 906
Db 221 ProfileAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240
Qy 907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCCACCCCGGCC 966
Db 241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260
Qy 967 GGCCTGAAGAAGAGAGAGCGTGACCGTGTGACGTGGCGCAGCGCTACTTTCAGCGTG 1026
Db 261 GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280
Qy 1027 CCCTCGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCATCAACAACGAG 1086
Db 281 ProLeuAspLysAspPheArgLysTrpAlaPheThrIleProSerIleAsnAsnGln 300
Qy 1087 ACCCGCGCATCGCTACCAAGTACACGTGTGCGCCAGAGGTGGAGGGCAGCGCCCGAC 1146
Db 301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320
Qy 1147 ATCTTCAGAGCAGATGACCAAGATCTCGAGCCCTTCGCGCGCCCGCAACCCCGAGATC 1206
Db 321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle 340
Qy 1207 GTGATCTACAG-----GCCCGCTGTAGTGGCGAGCGACCTGGAGATCGGCACGAC 1260
Db 341 IleIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360
Qy 1261 CGGCGCAAGATCGAGGAGCTGCGCAAGCATCTGTGCGCTGGGCTTCACCCACCCCGAC 1320
Db 361 ArgGluAlaValGluAspLeuAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380
Qy 1321 AAGAGCACCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1374
Db 381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400
Qy 1375 TGACCGCTGAGCCCATCGAGCTGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1434
Db 401 TrpThrValGlnProIleLysLeuProGluLysValTrpThrValAsnAspIleGln 420
Qy 1435 AAGCTGTGGCGCAAGCTGAATCGAGCTGCGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1494
Db 421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440
Qy 1495 CTGTGAAGTGTGTGCGCGCGCAAGGCCCTGACCGACATCGTGGCCCTCGACCGAGGAG 1554
```

Db 441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460  
 QY 1555 GCCGAGCTGGAGCTGGCGAGAACCGCGAGATCCTGCGGAGCGCGTGCACGGCGTGTAC 1614  
 Db 461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480  
 QY 1615 TACGACCCCGACGAGCACTGGTGGCCGAGATCCAGAAGCAGGGCCACGACCAGTGGACC 1674  
 Db 481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnTyrThr 500  
 QY 1675 TACGAGATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGC 1734  
 Db 501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520  
 QY 1735 ACCGGCCCAACCAACGAGCTGAAGCAGCTGACCGGAGCGGTGCAGAAGATGCCATGGAG 1794  
 Db 521 SerAlaHisThrAsnAspLysGlnLeuValGluValValArgLysValAlaThrGlu 540  
 QY 1795 AGCATCGTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAGGAGACCTGG 1854  
 Db 541 SerIleValIleTyrGlyLysThrProLysPheArgLeuProValGlnLysGluValTyr 560  
 QY 1855 GAGACCTGTGTGACCACTACTGGCAGGCCACTGTGATCCCGAGTGGAGTTCTGTAAAC 1914  
 Db 561 GluAlaTyrTyrThrAspHisTyrGlnAlaThrTyrPileProGluTyrGluPheValAsn 580  
 QY 1915 ACCCGCCCGCTGGTGAAGCTGTGTACGAGCTGGAGAGGAGGCCCATCATCGGCGCGAG 1974  
 Db 581 ThrProProLeuValLysLeuTyrTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600  
 QY 1975 ACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2034  
 Db 601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620  
 QY 2035 ACCGACCGGGCGCGCAGAGATCTGTGAGCTGTGACCGAGACCAACCAAGAGACCGAG 2094  
 Db 621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640  
 QY 2095 CTGCAAGGCATCAGCTGGCGCTGCGAGACGCGGAGGAGTGAACATCGTGACCGAC 2154  
 Db 641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660  
 QY 2155 AGCCAGTACGCCCTGGCATCATCCAGGCCCGACCGCCAGAGCGAGCGAGCGAGTGGTG 2214  
 Db 661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680  
 QY 2215 AACCAAGATCATCAGAGCTGATCAAGAGAGAGAGGTGTACTGAGCTGGTGGCCCGCC 2274  
 Db 681 SerGlnIleIleGluLeuIleLysLysGluArgValTyrLeuSerTyrValProAla 700  
 QY 2275 CACAGGGGCATCGCGGCAACGAGCATCGACAGCTGGTGAAGCAAGGGGCATCCGCAAG 2334  
 Db 701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720  
 QY 2335 GTGCTGTCTCTGAGCGGCATCGAT 2358  
 Db 721 IleLeuPheLeuAspGlyIleGlu 728

Search completed: June 2, 2005, 05:13:50  
 Job time : 106.917 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 03:13:23 ; Search time 34.0828 Seconds  
(without alignments)  
10815.332 Million cell updates/sec

Title: US-09-610-313B-30  
Perfect score: 4583  
Sequence: 1 gtcagccacacatggccga.....gggttagcaccggtaattc 2469

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09610313/runat 31052005 155137 15173/app query.fasta\_1.7893  
-DB=Issued Patents AA -QPMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1 1 101 @runat 31052005 155137 15173 -NCPU=6 -ICPU=3  
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pcp.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pcp.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pcp.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pcp.\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS.COMB.pcp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Query §			DB	ID	Description			
	Score	Match	Length						
1	3535	77.1	1005	2	US-07-743-357-1	Sequence 1, Appli			
2	3531.5	77.1	1003	2	US-07-743-357-9	Sequence 9, Appli			
3	3513	76.7	1015	3	US-08-463-210-9	Sequence 9, Appli			
4	3513	76.7	1015	3	US-09-124-900-3	Sequence 3, Appli			
5	3513	76.7	1015	4	US-08-463-028-9	Sequence 9, Appli			
6	3513	76.7	1016	2	US-07-743-357-2	Sequence 2, Appli			
7	3510	76.6	1003	4	US-09-309-572-17	Sequence 17, Appli			
8	3510	76.6	1003	4	US-09-718-096-17	Sequence 17, Appli			
9	3507	76.5	1016	2	US-07-743-357-5	Sequence 5, Appli			
10	3506	76.5	1016	2	US-07-743-357-3	Sequence 3, Appli			
11	3503	76.4	1016	2	US-07-743-357-4	Sequence 4, Appli			
12	3497.5	76.3	1003	2	US-07-743-357-10	Sequence 10, Appli			

13	3494.5	76.2	1003	2	US-07-743-357-8	Sequence 8, Appli			
14	3487	76.1	1004	2	US-07-743-357-7	Sequence 7, Appli			
15	3221.5	70.3	1014	4	US-09-319-588C-6	Sequence 6, Appli			
16	3179	69.4	913	2	US-07-743-357-22	Sequence 22, Appl			
17	3135.5	68.4	1350	4	US-09-952-060-35	Sequence 35, Appl			
18	3116	68.0	913	2	US-07-743-357-6	Sequence 6, Appli			
19	2860.5	62.4	875	4	US-09-952-060-6	Sequence 6, Appli			
20	2855	62.3	850	4	US-09-952-060-2	Sequence 2, Appli			
21	2827	61.7	562	3	US-09-117-217-14	Sequence 14, Appl			
22	2827	61.7	562	4	US-09-735-487-14	Sequence 14, Appl			
23	2814.5	61.4	875	4	US-09-952-060-8	Sequence 8, Appli			
24	2809	61.3	850	4	US-09-952-060-4	Sequence 4, Appli			
25	2796	61.0	560	4	US-10-205-641-1	Sequence 1, Appli			
26	2545.5	55.5	1018	4	US-09-206-551-46	Sequence 46, Appl			
27	2521.5	55.0	1016	4	US-09-625-972-24	Sequence 24, Appl			
28	2270	49.5	441	4	US-09-991-258-16	Sequence 16, Appl			
29	2247	49.0	1055	2	US-08-659-251-5	Sequence 5, Appli			
30	2247	49.0	1055	3	US-09-256-490-5	Sequence 5, Appli			
31	2247	49.0	1055	5	PCT-US96-11445-5	Sequence 5, Appli			
32	2228.5	48.6	3080	6	5223423-4	Patent No. 5223423			
33	2228.5	48.6	3080	6	5223423-4	Patent No. 5223423			
34	2192	47.8	427	3	US-09-690-265-1	Sequence 1, Appli			
35	1593	34.8	327	3	US-08-679-493A-71	Sequence 71, Appl			
36	1590	34.7	314	1	US-08-589-446-6	Sequence 6, Appli			
37	1590	34.7	314	1	US-08-444-882-6	Sequence 6, Appli			
38	1590	34.7	314	2	US-08-389-459A-6	Sequence 6, Appli			
39	1590	34.7	314	3	US-08-987-867A-6	Sequence 6, Appli			
40	1576	34.4	327	3	US-08-679-493A-72	Sequence 72, Appl			
41	1349	29.4	1150	3	US-09-238-303-9	Sequence 9, Appli			
42	1349	29.4	1150	4	US-09-946-239-9	Sequence 9, Appli			
43	1335	29.1	261	6	5320958-3	Patent No. 5320958			
44	1335	29.1	261	6	5320958-3	Patent No. 5320958			
45	1320	28.8	1031	3	US-08-811-682-15	Sequence 15, Appl			

ALIGNMENTS

RESULT 1

US-07-743-357-1  
; Sequence 1, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 1:





Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 Qy 1171 ATCTGGAGCCCTTCGCGCCCGCAACCCGAGATCGTGATCTACAGTACATGGACGAC 1230  
 Db 321 ILeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
 Qy 1231 CTGTACCTGGCGACGACCTCGAGATCGGCACCGCCCGCAAGATCGAGGAGCTGCGC 1290  
 Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuArg 360  
 Qy 1291 AAGCCTCTGCTGGCTGGGCTTCACCCCGCCGCAAGAACGACACGAGAGAGCCCGCC 1350  
 Db 361 GluHisLeuLeuLysTyrGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
 Qy 1351 TTCTGTGGATGGCTACGAGCTGCACCCGACAGTGGACCTGCGAGCCCATCGAGCTG 1410  
 Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
 Qy 1411 CCCGAGAAGGAGAGCTGGACCGTGCAACGACATCCAGAACTGGTGGCAAGCTGAACCTGG 1470  
 Db 401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
 Qy 1471 GCCAGCAGATCTACCCCGCGATCAAGTGGCGGAGCTGCGAAGCTGCGCGGCGCC 1530  
 Db 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLysLeuLysLeuArgGlyAla 440  
 Qy 1531 AAGCCCTGACGACATCGTCCCGCTGACCGAGAGCGCGAGCTGGAGCTGGCGGAGAAC 1590  
 Db 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460  
 Qy 1591 CGCGAGATCTCGCGGAGCCGCTGCGAGCGGTGCTACTACGACCCCGCAAGAGCACTGGT 1650  
 Db 461 ArgGluIleLysGlnGlyProValHisGlyValTyrTrpAspProSerLysAspLeuIle 480  
 Qy 1651 GCCGAGATCCAGAGAGCGCCAGCACAGTGGACCTTACAGATCTACAGAGCGCCCTTC 1710  
 Db 481 AlaGluIleGlnLysGlnGlyGlnGlyGlnTrpTrpTyrGlnIleTyrGlnGluGlnTyr 500  
 Qy 1711 AAGAACTGAAGACCGGACGAGTACGCAAGATCGCACCGCCCGCAACACGACGCTGAAG 1770  
 Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
 Qy 1771 CAGCTGACCGAGGCGCTGCAGAAAGATCCCATGAGAGCATCGTGATCTGGGGCAAGACC 1830  
 Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
 Qy 1831 CCAGAGTCCGCTGCGCCATCCAGAGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1890  
 Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
 Qy 1891 CAGGCCACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGG 1950  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
 Qy 1951 TACGAGCTGGAGAGGAGCCCATCATCGCGCCCGAGACCTTCTACGTGGACGCGCGGCC 2010  
 Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 Qy 2011 AACCGCAGACCAAGATCGGACGCGCGCTAGCTGACCGACCGCGCGCGCGCGAGATC 2070  
 Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 Qy 2071 GTGAGCTGACCGAGACCCACCAAGACGAGCTGCGAGCCATCCAGCTGGCCCTG 2130  
 Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGlnLeuGlnAlaIleHisLeuAlaLeu 640  
 Qy 2131 CAGGACACCGCGAGCGAGGTGAACATCGTACCGACGACGAGTACGCGCTGGGCATCATC 2190  
 Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 Qy 2191 CAGCCCGACCGCAAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCGAGCTGATC 2250  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680

Qy 2251 AAGAAGGAGAGGTGTACCTGAGCTGGTGCCTCCGCCCAAGAGGCGATCGCGCGCAACGAG 2310  
 Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
 Qy 2311 CAGATCGAAGAGTGTGTGAGCAAGGCGATCCGCAAGGTGCTGTCTCGAGCGCATCGAT 2370  
 Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720  
 RESULT 3  
 US-08-463-210-9  
 ; Sequence 9, Application US/08463210  
 ; Patent No. 6001977  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CHANG, Nancy T.  
 ; APPLICANT: GALLO, Robert C.  
 ; APPLICANT: WONG-STALL, Flossie  
 ; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA  
 ; NUMBER OF SEQUENCES: 11  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Morgan & Finnegan, L.L.P.  
 ; STREET: 345 Park Avenue  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10154-0053  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/463,210  
 ; FILING DATE: 05-JUN-1995  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/693,866  
 ; FILING DATE: 23-JAN-1985  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 06/659,339  
 ; FILING DATE: 10-OCT-1984  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Serunian, Leslie A.  
 ; REGISTRATION NUMBER: 35,353  
 ; REFERENCE/DOCKET NUMBER: 2026-4193US2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 758-4800  
 ; TELEFAX: (212) 751-6849  
 ; INFORMATION FOR SEQ ID NO: 9:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1015 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: HTLV-III  
 ; FEATURE:  
 ; NAME/KEY: Protein  
 ; LOCATION: 1..1015  
 ; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
 US-08-463-210-9  
 Alignment Scores: 8.35e-212 Length: 1015  
 Pred. No.: 3513.00 Matches: 658  
 Score: 3513.00  
 Percent Similarity: 94.27% Conservative: 33  
 Best Local Similarity: 89.77% Mismatches: 26  
 Query Match: 76.65% Indels: 16  
 DB: 3 Gaps: 3  
 US-09-610-313B-30 (1-2469) x US-08-463-210-9 (1-1015)

QY 220 TTCTTCCGCGAGGACCTGGCCCTTCCCGCAGGGCAAGGCCCGCGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 268 -----CCACAGCCAGCAGAACCGCGCCCAACAGCCCGCCACAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGlnThrArgAlaAsnSerProThrArg 40  
QY 304 CGCAGAGCTGCAGGTG-----CGCGCGCACAACCCCGCAGCGAGCGCGCGCGAGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
QY 358 CAGGGCAACCTG-----AATCTCCCGCAGATCACCTGTGGCAGCGCCCTCGTGAGC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
QY 412 ATCAAGGTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGCAGCACCGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
QY 472 CTGAGGAGATGAGCTCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531  
Db 101 LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGly 120  
QY 532 TTCTAAGGTGGCGCGCAGTACGACAGATCTGTGATCGAGATCTCGCGCAAGAGGCCATC 591  
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
QY 592 GGACACGTGTGTGCGCGCCACCGCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
QY 652 CTGGGCTGCACCTCGAATCTCCCATCAGCCCATCGAGACCGTGCCTCGTGAAGCTGAAG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
QY 712 CCGGCGATGAGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCC 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACCCCATCTCGAGGAGATCGAAGAGGCGCAAGATCACCAGATCGCGCGCGCAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGlnGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACAAACCCCGCTGTCGCCATCAAGAAGAGGACAGACCAAGTGGCGCAAG 891  
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCGCGAGCTGAACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGC 951  
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCCACCCTCGCGCTGMAAGAAGAGAGCGTGCACCGTGGAGCTGGCGCGCAC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCAGCTGCGCTCGACGAGACTTCGCGAAGTACACCGCTTCACCATCCCC 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACGAGACCCCGCGCATCCGTACAGTACAACTGCTGCCCGCAGGCTGG 1131  
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGCGAGCCCGACATCTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLys 340  
QY 1192 CGCAACCCCGAGATCGGTACTACAGTACATGACGACCTGTACGTGGCGCAGGACCTG 1251  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360

QY 1252 GAGATCGGCGAGCACCAGCCGCAAGATCGAGAGCTGGCAAGCACCTGCTGGCTGGGGC 1311  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1312 TTCAACACCCCGCAACAAGACACCAAGAGAGCCCTTCCTGTGTGATGGCTTACGAG 1371  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400  
QY 1372 CTGACCCCGCACAAGTGGACCGCTGACCCCATCGAGCTGCCGAGAGAGAGAGTGGACC 1431  
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1432 GTGAACACATCCAGAAGCTGGTGGCAAGCTGAACCTGGGCGCAGCAGATCTACCCGCGC 1491  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY 1492 ATCAAGGTGGCGCAGCTGTCAAGCTGTGGCGCGCGCAAGGCGCTGACGACATCGTG 1551  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1552 CCCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAACCCGCGAGATCTCTCGCGAGGCC 1611  
Db 461 ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1612 GTGCAACCGCTGTACTAGCACCCCGCAGCAGACCTGTGGCGCAGATCCAGACGAGGCC 1671  
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1672 CACGACAGTGGACTACACAGATCTACGAGAGCCCTTCAAGAACCTGAAACCGCGCAAG 1731  
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysProPheLysAsnLeuLysThrGlyLys 520  
QY 1732 TAGCCCAAGATGGCCGCCCCACCAACAGCAGCTGAAGCAGCTGACGAGCCGCTGCAG 1791  
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1792 AAGATCCCATGGAGACATCGTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCCATC 1851  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1852 CAGAAGGAGACCTGGGAGACCTGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAG 1911  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1912 TGGAGGTTCGTGAACACCCCGCTCGTGAAGCTGTGTACCACTGGAGAGAGAGGCC 1971  
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro 600  
QY 1972 ATCATCGCGCGCAGAGACTTCTAGCTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGC 2031  
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2032 AAGSCCGCTACGTGACCGAGCCGCGCGCGCAGAGATCTGAGCCTGACCGAGCACACC 2091  
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2092 AACCAAGACCGAGCTGACAGGCCATCCAGCTGCGCTCGCAGGACAGCGCGCAGGAGTG 2151  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2152 AACATCGTGACGACAGCAGTACGCCCTGGGCGATCATCAGGCCCGCCCGCAGAGGCC 2211  
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2212 GAGAGCGAGCTGTGACACAGATCATCGACAGCTGATCAAGAGGAGAGAGTGTACTGT 2271  
Db 681 GluSerGlnLeuValAsnGlnIleIleGlnLeuIleLysLysValLysValTyrLeu 700  
QY 2272 AGCTGGTGGCCCGCCCAAGAGGCGCATCGCGCGCAACAGCAGAGATCGCAAGCTCGTGCAGC 2331  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyLysGlnGluGlnValAspLysLeuValSer 720  
QY 2332 AAGGGCATCCGCAAGGTGTGTTCTTCTGGACGGCATCGAT 2370



```
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
|||||
RESULT 4
US-09-124-900-3
; Sequence 3, Application US/09124900
; Patent No. 6268484
; GENERAL INFORMATION:
; APPLICANT: KATINGER, Hermann
; APPLICANT: BUCHACHER, Andrea
; APPLICANT: ERNST, Wolfgang
; APPLICANT: BALLAUN, Claudia
; APPLICANT: PURTSCHER, Martin
; APPLICANT: TRKOLA, Alexandra
; APPLICANT: PREDL, Renate
; APPLICANT: SCHMATZ, Christine
; APPLICANT: KLIMA, Annelies
; APPLICANT: STEINDL, Franz
; APPLICANT: MUSTER, Thomas
; TITLE OF INVENTION: HIV-Vaccines
; FILE REFERENCE: 1939-112P
; CURRENT APPLICATION NUMBER: US/09/124,900
; CURRENT FILING DATE: 1998-07-30
; PRIOR APPLICATION NUMBER: PCT/EP95/01481
; PRIOR FILING DATE: 1995-04-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-124-900-3

Alignment Scores:
Pred. No.: 8,35e-212 Length: 1015
Score: 3513.00 Matches: 658
Percent Similarity: 94.27% Conservative: 33
Best Local Similarity: 89.77% Mismatches: 26
Query Match: 76.65% Indels: 16
DB: 3 Gaps: 3

US-09-610-313b-30 (1-2469) x US-09-124-900-3 (1-1015)
Qy 220 TTCTTCGGCAGGACCTGGGCTTCCCGCAGGGAAGGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyIleArgGluPheSerSerGluGln 20
Qy 268 -----CCGAGCGAGCAGAACCGCGCCCAACAGCCCGCCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
Qy 304 CGCGAGCTGAGGTG-----CGGGCGACACCCCGCAGCAGCGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyIleArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
Qy 358 CAGGGCACCTCG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
Qy 412 ATCAAGTGGCGCGCCAGATCAAGAGGCGCCCTGCTGGACACCGCGCGCGCAGCACCGGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
Qy 472 CTGGAGGAGATGAGCTGCGCGCAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyIleArgTrpLysProLysMetIleGlyGlyIleGlyGly 120
Qy 532 TTCAATCAAGTGGCGCAGTACGACCATCTGATCGAGATCTCGCGCAAGAGGCGCATC 591
Db 121 PheIleLysValArgGlnIleThrAspGlnIleLeuIleGlyIleCysGlyHisLysAlaIle 140
Qy 592 GGCACCGTGTGATCGGCGCCCGCCCGCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651
|||||
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
Qy 652 CTGGGCTGCACCTGAACCTTCCCATCGACGCCCATCGACCGCTGCCCGCTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
Qy 712 CCGGCGATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTACCAGGAGAGATCAAGAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
Qy 772 CTGACCGCCATCTCGAGGAGATGGAAGAGAGGAGGCAAGATCACCAAGATCGGCCCGG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGlu 220
Qy 832 AACCCCTACAACACCCCGCTTCCGCATCAAGAAGAAGACAGACCAAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240
Qy 892 CTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTCTTGGGAGGTGCAGCTGGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
Qy 952 ATCCCGCACCCCGCGCTGAAGAAGAAGAGAGCGTGACCGTGTGACGCTGGCGGCGAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280
Qy 1012 GCCTACTTCAGCGTGCCTTGGACGAGGACTTCCGCAAGTACACCGCTTCCACATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
Qy 1072 AGCATCAACAACAGAGACCCCGCGCATCCGCTACAGTACAAACGTGTGCCCGAGGCTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
Qy 1132 AGGGGAGCGCCAGCATCTTCAGACGACGATGACCAAGATCTCGAGCGCTTCCGGCGC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
Qy 1192 CGCAACCCCGAGATCGTGTACACAGTACATGACGACCTGTACCTGGCGACGACCTG 1251
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeu 360
Qy 1252 GAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTCGCAAGCACCTGTGCTGGCGG 1311
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuAspTrpGly 380
Qy 1312 TTCACACCCCGCAGCAAGAAGCACGACGACCTGTACCTGTGGTGGCTAGCGAG 1371
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400
Qy 1372 CTCACCCCGCAGCAAGTGGACCGCTGCGAGCCCATCGAGCTGCCCGCAGAGAGAGCTG 1431
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
Qy 1432 GTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCCAGCGCAGATCTACCCGCG 1491
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
Qy 1492 ATCAAGTGGCGCGCATGTGCAAGCTGTGCGCGCGCGCCAGCGCTTCCGCGAGCGCC 1551
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
Qy 1552 CCCTCAGCAGGAGCGCGAGCTGGAGCTGGCGCGCGCGCCAGCGCTTCCGCGAGCGCC 1611
Db 461 ProLeuThrGluGluAlaGluLeuLeuAlaGluAsnArgGluLeuLysGluPro 480
Qy 1612 GTGCACGCGGTGTACTACGACCCCAAGAGCCTGTGTGGCCGAGATCTCAGAGACGAGGC 1671
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
Qy 1672 CAGCAGCAGTGGACCTTACAGATCTACAGGAGCGCTTCAAGACCTGAGACCGCGCAAG 1731
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
```

```

QY 1732 TACGCCAAGATCCGACCGCCGCCACCAACGAGCTGAGACGAGCTGACCGAGCGCTGCAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1792 AAGATCGCCATCGAGAGCATCGTATCTGGGCAAGACCCCAAGTCCGCTGCCCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProLeu 560
QY 1852 CAGAAGAGACTGGGACACTGTGTGACCGACTACTGCGAGCCACCTGGATCCCGAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluThrTrpGlnAlaThrTripleProGlu 580
QY 1912 TGGAGTTCTGTGAACACCCCGCCCTGTGTGAGCTGTGTGTACCTGAGAGAGAGGCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1972 ATCATCGCGCGGACGACCTTCTACGTGTGACGCGCGCCCAACCGGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2032 AAGCGCGCTACGTGACCGACCGCGCGCGGAGAGATCGTGAGCTGACCGAGACACCC 2091
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2092 AACCAAGAGCGAGCTGACGCCATCCAGTGGCCCTGCGAGGACGAGCGGAGGAGGTG 2151
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2152 AACATCGTGACCGACGACGAGTACGCTGGGATCATCCAGGCGCGGACCGAGAGGAC 2211
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2212 CAGAGCGAGCTGTGTGAACACGATCATCCAGGAGCTGTGTGTGAAGAGAGAGTGTACCTG 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2272 AGCTGGGTGCGGCCCAAGGCGATCGCGGCGGACGAGAGATCGACAAGCTGTGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2332 AAGGCGATCCGCAAGGTCTGCTTCTGCGAGCGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

```

RESULT 5

```

US-08-463-028-9
; Sequence 9, Application US/08463028
; Patent No. 6610476
; GENERAL INFORMATION:
; APPLICANT: CHANG, Nancy T.
; APPLICANT: GALLO, Robert C.
; APPLICANT: WONG-STAAAL, Flossie
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HTLV-III DNA
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESS: Morgan & Finnegan, L.L.P.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,028
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/693,866

```

```

; FILING DATE: 23-JAN-1985
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/659,339
; FILING DATE: 10-OCT-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4193US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1015 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: HTLV-III
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1015
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
; US-08-463-028-9

```

```

Alignment Scores:
Pred. No.: 8,35e-212 Length: 1015
Score: 3513.00 Matches: 658
Percent Similarity: 94.27% Conservative: 33
Best Local Similarity: 89.77% Mismatches: 26
Query Match: 76.65% Indels: 16
DB: 4 Gaps: 3

```

US-09-610-313B-30 (1-2469) x US-08-463-028-9 (1-1015)

```

QY 220 TTCCTCCGCGAGGACCTGGCTTCCCGAGGCAAGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 268 -----CCGAGCGAGGAGAACCGCGCCCAACAGCCCGCCACCGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGCGCACAAACCCCGAGCGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCGACCCCTG-----AACTTCCCGAGATCACCTGTGGCAGCGCCCTGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGTGGCGCGCGAGATCAAGAGGCGCTGTGGACACCGCGCGCGCGCGCGCGCGC 471
Db 81 IleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal 100
QY 472 CTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGGATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY 532 TTCATCAAGGTGCGCCAGTACGACAGATCTGTGATCGAGATCTCGCGCAAGAGCCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACTTCCCATCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGC 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180

```

QY 712 CCCGCCATGACGCCCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCC 771  
DB 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
QY 772 CTGACGCCCATCTGCGAGGAGATGGAAGAGGGCAAGATCAACCAAGATCGGCCCGAG 831  
DB 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
QY 832 AACCCCTACAACACCCCGCTTCCCATCAAGNAGAGACACCAAGTGGCGCAAG 891  
DB 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGTGTGACTTCCCGAGCTGAACAAGCGCACCCAGACTTCTGGAGGTGCAGCTGGC 951  
DB 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCCACCCCGCCCTGAAGAAGAAGAGCGTGACCGTGTGGACGTGGCGCAC 1011  
DB 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCAGCTGCCCTGGACGAGACTTCCGCAAGTACACCGCCTTCACCATCCC 1071  
DB 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
QY 1072 AGCATCAACACGAGACCCCGGCATCCGCTACCAAGTACCAACGTGCTGCCCGCCAGGGCTGG 1131  
DB 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGCAGCCCGACATCTTCCAGAGCAGATGATGACCAAGATCTGTGGAGCCCTTCCGGGCC 1191  
DB 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGNACCCCGAGATCGTACTACCACTACATGACGACCTGTACGTGGCGACGACCTG 1251  
DB 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360  
QY 1252 GAGATCGGCAGACCCCGCCCAAGATCGAGAGCTGGCGCAAGCACCTGCTGCGCTGGGGC 1311  
DB 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1312 TTCAACCCCGCAAGAACAGCACCAAGAGAGCCCTTCTGTGTGGATGGGCTACGAG 1371  
DB 381 LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu 400  
QY 1372 CTGCACCCCGACAGTGGCCGTGCGCCCATCGAGCTGCCCGCCAGAGGAGAGCTGGACC 1431  
DB 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1432 GTGAACGACATCCAGAGCTGGTGGGCAAGCTGAATCGGCGCCAGCATCTACCCCGGC 1491  
DB 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
QY 1492 ATCAAGTGGCCGACGTGTGCAGCTGCTGCGGCGCCCAAGGCCCTGACCGACATCGTG 1551  
DB 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1552 CCCTGTACCGAGGCGCGAGCTGGAGCTGCGCGAGACCGAGATCTCTCGCGGAGCCC 1611  
DB 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1612 GTGCACCGCGTGTACTACGACCCCGACAGACCTGGTGGCCGAGATCCAGAGACGGGC 1671  
DB 481 ValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1672 CACGACGAGTGACCTTACCAAGATCTACAGAGCCCTTCAAGAACCTTGAAGACGGCAAG 1731  
DB 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1732 TAGCCCAAGATGCCCGCCCGCCACCAAGCAGCTGAAGCAGCTGACCGAGCGCGTGCAG 1791  
DB 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
QY 1792 AAGATCGCCATGGAGATCGTGTATCTGGGGCAAGACCCCAAGTTCCCGCTGCCCATC 1851

DB 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1852 CACAAGAGACCTGGAGAGACCTGGTGGACCGACTACTGGCAGCGCCACCTGGATCCCCGAG 1911  
DB 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1912 TGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCCC 1971  
DB 581 TrpGluPheValAsnThrProProLeuValLysLeuThrTrpTyrGlnLeuGluLysGluPro 600  
QY 1972 ATCATCGGCGCCGAGACCTTCTACGTGGAGCGCGCCCAACCCGAGACCAAGATCGGC 2031  
DB 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2032 AAGCCCGGTACTGTAGACCGACCGCGCGCGCAGCAAGATCGTGGAGCTGACCGAGACCACC 2091  
DB 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2092 AACCAAGAAGACCGAGCTGCAGGCGCATCCAGCTGGGCATCATCCAGCCCGCAGCGCAGCGGTG 2151  
DB 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2152 AACATCGTGACCGACAGCCAGTACGCGCTGGGCATCATCCAGCCCGCAGCGCCGCAAGAGC 2211  
DB 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2212 GACAGCGAGCTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTG 2271  
DB 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2272 AGCTGGTGGCCCGCCCAAGGCGCATCGCGGCAAGCAGCAGATCGCAAGCTGGTGAGC 2331  
DB 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720  
QY 2332 AAGGCGATCGCAAGTGTCTTCTTCTGGAGCGGCATCGAT 2370  
DB 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 6  
US-07-743-357-2  
; Sequence 2, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900



```

Db      601  ILeValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY      2032  AAGCCGGCTACGTGACCGACCGGGCCGCGAGATCGTACGCTGACCGAGACACC 2091
Db      621  LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY      2092  AACCAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGACAGCGCAGCGAGGTG 2151
Db      641  AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY      2152  AACATCGTGACCGACAGCCAGTAGTCGCCCTGGCATCATCCAGGCCCGACCGCAAGAGC 2211
Db      661  AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY      2212  GAGAGCGAGCTGGTGAACACAGATCATCGACAGCTGATCAAGAGGAGAGGTGTACCTG 2271
Db      681  GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY      2272  AGCTGGGTGCCCGCCCAAGGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
Db      701  AlaIrpValProAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720
QY      2332  AAGGGCATCCGCAAGTGTGTTCTCGACGGCATCGAT 2370
Db      721  AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7
US-09-309-572-17
; Sequence 17, Application US/09309572
; Patent No. 6440730
; GENERAL INFORMATION:
; APPLICANT: Heinrich-Pette-Institut
; TITLE OF INVENTION: Retroviral hybrid vectors pseudotyped with LOMV
; FILE REFERENCE: P50489
; CURRENT APPLICATION NUMBER: US/09/309,572
; CURRENT FILING DATE: 1999-05-11
; EARLIER APPLICATION NUMBER: DE 198 56 463
; EARLIER FILING DATE: 1998-11-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-309-572-17

Alignment Scores:
Pred. No.:      1,28e-211      Length:      1003
Score:          3510.00      Matches:      656
Percent Similarity: 95.70%      Conservative: 34
Best Local Similarity: 90.98%      Mismatches: 27
Query Match:     76.59%      Indels:      4
DB:              4              Gaps:        2

US-09-610-313B-30 (1-2469) x US-09-309-572-17 (1-1003)
QY      220  TTCTTCGCGAGGACCTGGCTTCCCGCCAGGCGAAGCCCGCGAGTTCCCGAGCGAGCAG 279
Db      1   PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY      280  AACCGCGCCAAACGCCCCACAGCCGCGAGCTGCAGGTG-----CGCGCGCAACACCC 333
Db      21  ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY      334  CGCAGCGCAGCGCGCGCGAGCGCCAGCGCACCGACCTG-----AACTTCCCGCATCAC 387
Db      41  LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY      388  CTGTGGCAGCGCCCGCTGGTGAGCATCAAGTGGCGCGCGCGAGATCAAGAGGCGCCTGCTG 447

```

```

Db      61  LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY      448  GACACCGCGCGCCAGCACACCGTGTGCGAGGAGATGAGCTGCCCGCGAAGTGAAGCC 507
Db      81  AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY      508  AAGATGATCGCGCGCATCGCGCGCTTCAACAGGTGCGCGAGTACGACCAAGATCCTGATC 567
Db      101  LysMetIleGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120
QY      568  GAGATCTGGCGCAAGAGCCATCGGCACCGTCTGATCGGCCCGCCACCGCTGAACATC 627
Db      121  GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY      628  ATCGCGCGCAACATGCTGACCCAGCTGGGTGCACCTGAACTTCCCATCAGCCCGCATC 687
Db      141  IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY      688  GAGACCGTGCCTGAAAGCTGAAGCCCGCGCATCGACGGCCCGCCCAAGGTGAAGCAAGT 747
Db      161  GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY      748  CTGACCGAGAGAAATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGGAAGAGGGCG 807
Db      181  LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY      808  AGATCACCAAGATCGCGCGAGAACCCCTACAACACCCCGCTGTTCGCATCAAGAAG 867
Db      201  LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY      868  AAGCAGACGACCAAGTGGCGCAAGCTGTGTGAGCTTCCGCGAGCTGAACAGCGCACCC 927
Db      221  LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY      928  GACTTCTGGAGAGTGCAGCTGGGCATCCCGCCACCGCGCGCGCTGAAGAAGAAGAGAGC 987
Db      241  AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysSer 260
QY      988  GTGACCGTGTGACGTGGCGGACGCTACTTCCAGGTGCGCGCTGACGAGGAGCTTCCGC 1047
Db      261  ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY      1048  AAGTACACCGCTTCCACCATCCCGCATCAACAACGAGACCCCGCGCATCCGCTACCAG 1107
Db      281  LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY      1108  TACAAGTGTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTCCAGAGCAGCATGACC 1167
Db      301  TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThr 320
QY      1168  AAGATCTGGAGCCCTTCCGCGCGCCGCAACCCCGAGATCGTGTATACCAAGTACATGGAC 1227
Db      321  LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY      1228  GACTGTGTACGTGGCGCAGCACCTGGAGATCGGCAGCACCGCCCGCAAGATCGAGGAGCTG 1287
Db      341  AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY      1288  CGCAAGACCTGTGCTGGCTGGGCTTCAACACCCCGCAAGAAGACCAAGAGAGGAGCC 1347
Db      361  ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY      1348  CCTCTCTGTGATGGCTACGAGCTGCACCCGCAAGTGGACCGTGCAGCCCATCGAG 1407
Db      381  ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY      1408  CTGCCCGCAGAGGAGCTGGACCGCTGAACACATCCAGAAAGCTGTGGCGCAAGCTGAAC 1467
Db      401  LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY      1468  TGGCGCAGCCAGATCTTACCCCGCGCATCAAGTGGCGCGCGAGCTGTGCAAGCTGCTG 1527
Db      421  TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440

```

```
QY 1528 GCCAAGCCCTGACCGACATCGTCCCTGACCGAGGCGCGAGCTGGAGCTGGCCGAG 1587
Db 441 ThrLysAlaLeuThrGluValProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1588 AACCGCAGATCCCGCGGCGCGTGCACGGCTGTACTACGACCCCGACGAGCCTG 1647
Db 461 AsnArgGluLeuLeuGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1648 GTGCCCGAGATCCAGAGCAGGCGCACGACAGTGGACCTACAGATCTACGAGAGCCC 1707
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
QY 1708 TTCAGAACCTGAAGACCGCAAGTACGCCAAGATGCGCACCGCCACCAACAGCGTG 1767
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGCGCGTGCAGAAGATGCCATGAGAGCATCGTGTGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTyrGlyLys 540
QY 1828 ACCCCCAAGTCCCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC 1887
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTyrGluAlaTyrThrGluTyr 560
QY 1888 TGGCAGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTG 1947
Db 561 TrpGlnAlaThrTyrIleProGluTyrPheValAsnThrProProLeuValLysLeu 580
QY 1948 TGGTACCAGCTGGAGAGGAGCCCATCATCGGCGCGGAGACCTTCTACGTGGAGCGCC 2007
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGCGCGCAAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620
QY 2068 ATCGTGAAGTCCCGCAGCACCAACCAAGACCGAGTGCAGGCGCATCCAGCTGGCC 2127
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2128 CTGCAGACAGCGCAGCGAGGTGAACATCGTGACCGACGACGACGCTGGGCGATC 2187
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCCAGCCCGCACAGAGCGAGCGAGCTGGTGAACACGATCATCGACAGCTG 2247
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGGAGAGGTGTACCTGAGTGGGTGCGCCGCCCAAGGGCATCGCGCGCAAC 2307
Db 681 IleLysLysGluLysValTyrLeuAlaTyrValProAlaHisLysGlyIleGlyAsn 700
QY 2308 GAGCAGATCGACAGCTGGTGAGCAAGGGCATCGCGAGGTGCTGCTTCCTGGAGCGCATC 2367
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db 721 Asp 721
```

## RESULT 8

```
US-09-718-096-17
; Sequence 17, Application US/09718096
; Patent No. 6589763
; GENERAL INFORMATION:
; APPLICANT: Von Laer, Meike-Dorothee
; TITLE OF INVENTION: RETROVIRAL HYBRID VECTORS PSEUDOTYPED WITH LCMV
; FILE REFERENCE: 35-195
; CURRENT APPLICATION NUMBER: US/09/718,096
; CURRENT FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: DE 19856463.5
; PRIOR FILING DATE: 1998-11-26
```

```
; PRIOR APPLICATION NUMBER: EP 99250415.9
; PRIOR FILING DATE: 1999-11-25
; PRIOR APPLICATION NUMBER: US 09/309,572
; PRIOR FILING DATE: 1999-05-11
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
; FEATURE:
; OTHER INFORMATION: pol polyprotein
US-09-718-096-17
```

```
Alignment Scores:
Pred. No.: 1,28e-211 Length: 1003
Score: 3510.00 Matches: 656
Percent Similarity: 95.70% Conservative: 34
Best Local Similarity: 90.98% Mismatches: 27
Query Match: 76.59% Indels: 4
DB: 4 Gaps: 2
```

US-09-610-313B-30 (1-2469) x US-09-718-096-17 (1-1003)

```
QY 220 TTCTTCGCGAGGACCTGGCGCTTCCCCAGGGCAAGCGCCGCGAGTTCCCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCCCAACAGCCCGCCAGCGCGAGCTGCAGGTG-----CGCGGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCGAGCGCACCCCTG-----AACTTCCCCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCGCTGGTGCAGCATCAAGTGGCGCGCGCAGATCAAGAGCGCCCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGACACGCGTGTGAGGAGATGAGCTGCCCGGCAAGTGAAGGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCCAGTACGACCATCTGTATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValGlyGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGGCGCAAGAGGCGCATCGGCACCGTGTGTATCGGCCCGCCCGCTGAAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATCGTGCACCGCGCTGACCTGACCTGAACCTTCCCATCAGCCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGGCCGTGAAGCTGAAGCGCGCATGGACGCGCCCGCCCAAGGTGAAGCAGTGG 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCGCTGACCGCCCATCTGCGAGGAGATGGAGAGAGGCG 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCGCGAGAACCCCTTACAAACACCCCGCTTCGCCCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCAGCAAGTGGCGCGCAGCTGGTGAACCTTCCCGCAGCTGAACAACGCGCAC 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGGAGGTGACGTGGGCATCCCCCACCCCGCGCGCTGAAGAAGAAGAGAGC 987
```

Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysSer 260  
QY 988 GTACCGTGTGACGTGGCGAGCCCTACTTACGTGTCGCCCTTGACGAGGACTTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCGCTTCACCATCCCGCAGCATCAACAAGCAGACCCCGCATCCGCTACCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnGlnThrProGlyIleArgTyrGln 300  
QY 1108 TACAACGTGTGCTGCCCGCAGGCTGGAAGGCGAGCCCGCAGCATCTTCACAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnCysSerMetThr 320  
QY 1168 AAGATCTGTGAGCGCTTCCCGCGCGCAACCCCGAGATCGTGTATCTACCATGTCATGGAC 1227  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
QY 1228 GACCTGTACGTGGCGAGCATCGGATCGGCAGCAGCCCGCCCAAGATCGAGGAGCTG 1287  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1288 CGCAAGCACCTGTGCTGCGCTGGGGCTTCACCAACCCCGCAGCAAGACACCAAGAGAGGCC 1347  
Db 361 ArgGlnHisLeuLeuArgTyrGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1348 CCTTCTGTGTGGTACGAGTGCACCCGACCAAGTGGACCGTGGACCCATCGAG 1407  
Db 381 ProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1408 CTCCCGAGAGAGAGCTGCAGCGTGAACAGCATCCAGAGCTGTGGCGCAAGCTGAAC 1467  
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1468 TGGCGCCAGCAGATCTACCCCGCGCATCAAGGTGGCGCGAGCTGTGCAAGCTGTGCGCGGC 1527  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
QY 1528 GCGAAGCCCTGACCGACATCGTGCCTCCCTGACGAGAGCGCGAGCTGGCGCGAG 1587  
Db 441 ThrLysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1588 AACCGCAGATCTTCGCGAGCGCGTGCACGGGTGCTACTACGACCCCGCAGAGGACCTG 1647  
Db 461 AsnArgGlnIleLysGluProValHisGlyValTyrAspProSerLysAspLeu 480  
QY 1648 GTGCGCCAGATCCAGAGAGGCGCACGACAGTGGACCTTACAGATCTACAGAGGCC 1707  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500  
QY 1708 TTCAAGAACTGAAGCCGCAAGTACGCCAAGATGGCACCGCCCGCACCAACAGCAGCTG 1767  
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspVal 520  
QY 1768 AAGCAGCTGACCGAGCGCTGCAGAGATCGCATCGAGAGCATCTGTGATCTGGGCAAG 1827  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
QY 1828 ACCCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTTGGGAGACCTGTGGACCGACTAC 1887  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTyrTrpThrGluTyr 560  
QY 1888 TGGCAGCCACCTGGATCCCGAGTGGGAGTTGTGAACACCCCGCCCTGGTGAAGCTG 1947  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
QY 1948 TGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGCC 2007  
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600  
QY 2008 GCCAACCGCAGAGCAAGATCGGAGCGCGGTGTACGTGACCGAGCGGCGCGCGAGAG 2067

Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620  
QY 2068 ATCGTGAGCTGACCCAGAGACCAACACAGAGCTGACAGCGCATCCAGCTGGCC 2127  
Db 621 ValValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
QY 2128 CTGACGAGCAGCGCGCAGCGAGGTGAACATCGTGACCGCAGCAGCGCTGGCGCATC 2187  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
QY 2188 ATCCAGGCCAGCCCGCAGCAGAGCAGAGCTGGTGAAACCATCATCATCGAGCAGCTG 2247  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
QY 2248 ATCAAGAGGAGAGGTGTACTGAGCTGGTGGTCCCGCCACCAAGGCGCATCGCGGCAAC 2307  
Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
QY 2308 GAGCAGATCGACAAGCTGGTGAAGAGGCGCATCCGCAAGGTGCTGTTCTTGGAGCGCATC 2367  
Db 701 GluGlnValAspGlyLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2368 GAT 2370  
Db 721 Asp 721  
RESULT 9  
US-07-743-357-5  
; Sequence 5, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1016 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEetical: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: BRU



## US-07-743-357-5

## Alignment Scores:

Pred. No.: 1.98e-211 Length: 1016  
Score: 3507.00 Matches: 656  
Percent Similarity: 94.41% Conservative: 36  
Best Local Similarity: 89.50% Mismatches: 25  
Query Match: 76.52% Indels: 16  
DB: 2 Gaps: 3

US-09-610-313B-30 (1-2469) x US-07-743-357-5 (1-1016)

Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGGCAAGGCCCGGAGTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
Qy 268 -----CCCGAGCGAGAGAACCGCGCAACAGAGCCCGCAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40  
Qy 304 CGCGAGCTGCAGGTG-----CGCGGGCAACCCCGCAGCGAGCGCGCGCGCGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArg 60  
Qy 358 CAGGGCACCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
Qy 412 ATCAAGTGGCGCGCCAGATCAAGGAGCCCTGTGGACACCGCGCGCGCGCACACCGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
Qy 472 CTGAGGAGATGAGCTTCCCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGC 531  
Db 101 LeuGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyGly 120  
Qy 532 TTGATCAAGTGGCGCGCCAGTACAGATCTGTGATCGAGATCTCGCGCAAGAGCCATC 591  
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140  
Qy 592 GGACCGTGTGATCGCGCCCGCACCCCGTGAACATCATCGCGCGCAACATGCTCACCCAG 651  
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160  
Qy 652 CTGGGCTGCACCTTGAATCTCCCATCAGCCCGCATCAGACCGTGGCGCGTGAAGCTGAAG 711  
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180  
Qy 712 CCCGGCATGAGCGCCCGCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCC 771  
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200  
Qy 772 CTGACCGCCATCTCGGAGGAGATCGAGAGAGGCGCAAGATCACCAAGATCGCGCCGAG 831  
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220  
Qy 832 AACCCCTTACACACCCCGTGTCCCATCAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891  
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
Qy 892 CTGTGTGACTTCCCGAGCTGAAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGC 951  
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
Qy 952 ATCCCCCACCCCGCGCTGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
Qy 1012 GCCTACTTTCAGCTGCGCTGGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCC 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300  
Qy 1072 AGCATCAACACGAGACCCCGCGCATCCGCTACAGTACACAGTGTGCCCCAGGGCTGG 1131

Db 301 SerIleAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
Qy 1132 AAGGCGAGCCCGCATCTTCCAGAGCAGATGACCAAGATCTCGAGCCCTTCGCGGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys 340  
Qy 1192 CGCAACCCCGAGATCGTGATCTACAGTACATGACGACCTGTGTGGGCGAGGACCTG 1251  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeu 360  
Qy 1252 GAGATCGCGCGAGCAGCGCCCAAGATCGAGAGCTGCGCAAGACCTGCTGGCTGGGGC 1311  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
Qy 1312 TTCAACACCCCGCAACAAGAGCAGAGAGCGCCCTTCTGTGTGATGGGTACTACGAG 1371  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400  
Qy 1372 CTGCACCCCGCAACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGTGGACC 1431  
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
Qy 1432 GTGAACGACATCCAGAGCTGTGGCGCAAGCTGAACCTGGCGCAGCCAGATCTACCCCGC 1491  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440  
Qy 1492 ATCAAGTGGCGCGAGCTGTGCAAGCTGTGGCGCGCGCAAGGCGCTGACGACATCGTG 1551  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
Qy 1552 CCCGTGACCGAGGAGCGAGCTGTGAGTGGCCGAGAACCGCGAGATCTCTCGCGAGGCC 1611  
Db 461 ProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
Qy 1612 GTGCACCGCTGTACTAGACCCCGCAGCAAGCTGTGGCGCAGAGTCCAGAGCAGCGC 1671  
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
Qy 1672 CACGACAGTGGAGCTTACCAGATCTACGAGAGCGCTTCAAGAACCTTGAACACCGCGCAG 1731  
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnLysProPheLysAsnLeuLysThrGlyLys 520  
Qy 1732 TAGCCCAAGATGGCCACCGCCCAACCAAGCTGAAGAGAGTGAAGCGCGCGTGCAG 1791  
Db 521 TyrAlaArgThrArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540  
Qy 1792 AAGATCCCATGGAGACATCGTGTGCTGGGGCAAGACCCCAAGTTCCTGCTGCCATC 1851  
Db 541 LysIleThrThrGluSerIleValIleTrpLysThrProLysPheLysLeuProIle 560  
Qy 1852 CAGAGGAGACCTGGGAGACCTGTGTGACGAGTACTTGGAGAGCGCACCTGGATCCCGAG 1911  
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
Qy 1912 TGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGTGATCCAGCTGGAGAGAGAGGCC 1971  
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro 600  
Qy 1972 ATCATCGCGCGGAGACTTCTAGCTGACCGCGCGCGCCCAACCGCGAGACCAAGATCGGC 2031  
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaSerArgGluThrLysLeuGly 620  
Qy 2032 AAGGCGCGCTAGCTGAGCCGCGCGCGCGCGCAGAGAGTCTGTGACCTGACCGAGACCAAC 2091  
Db 621 LysAlaGlyTyrLeuThrAsnLysGlyArgGlnLysValValThrLeuThrAspThrThr 640  
Qy 2092 AACCAAGAGCCGAGTGCAGGCCATCCAGCTGCCCTGCAAGACAGCGCGAGGAGTG 2151  
Db 641 AsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
Qy 2152 AACATCTGACCGACCGCAGTACCGCTTGGGATCATCCAGGCCCGCGCAGCAAGGC 2211

Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
 Qy 2212 GAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAGAAGGTGTACCTG 2271  
 Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGlyValTyrLeu 700  
 Qy 2272 AGCTGGGTGCCCGCCACAGAGGGATCGGGGGCAACGAGCAGATCGACACAGCTGGTGAGC 2331  
 Db 701 AlaTyrValProAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720  
 Qy 2332 AAGGGCATCGCAAGGTGCTTCTTCGACGGCATCGAT 2370  
 Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733  
  
 RESULT 10  
 US-07-743-357-3  
 ; Sequence 3, Application US/07743357  
 ; Patent No. 5858646  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kang, Yong C.  
 ; TITLE OF INVENTION: Polypeptide having immunological  
 ; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
 ; NUMBER OF SEQUENCES: 22  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: KIRBY EADES GALE BAKER  
 ; STREET: Box 3432, Station D  
 ; CITY: Ottawa  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: K1M 1H8  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/743,357  
 ; FILING DATE: 21-AUG-1991  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/CA90/00062  
 ; FILING DATE: 23-FEB-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Gale, Edwin J.  
 ; REGISTRATION NUMBER: 28,584  
 ; REFERENCE/DOCKET NUMBER: 30924-2  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (613) 237-6900  
 ; TELEFAX: (613) 237-0045  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1016 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: not relevant  
 ; MOLECULE TYPE: protein  
 ; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: internal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Human immunodeficiency virus type 1  
 ; STRAIN: BH5  
 US-07-743-357-3

[illegible]

US-09-610-313B-30 (1-2469) x US-07-743-357-3 (1-1016)

Qy	220	TTCTTCGGCAGACCTGGCTTCCCCAGGGCAAGCCGCGAGTTC-----	267
Db	1	PhcPheArgGluAspLeuAlaPheLeuGlnGlyLeAlaArgGluPheSerSerGluGln	20
Qy	268	-----CCACGCGACAGAACCGCGCCCAACAGCCACCCACCCAGC	303
Db	21	ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg	40
Qy	304	CGCGAGCTGCAGGTG-----CGCGCGCAACACCCCGCAGCGAGCGCCGCGCAGCGC	357
Db	41	ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg	60
Qy	358	CAGGCAACCTG-----AACTTCCCCAGATCACCTGTGGCAGCGCCCTCGGTGAGC	411
Db	61	GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValTrp	80
Qy	412	ATCAAGTGGCGGCAGATCAAGGAGGCCCTCTCGACACCGCGCGCAGACACCGCTG	471
Db	81	IleuysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal	100
Qy	472	CTGAGAGATGAGCCTCGCCGCAAGTGAAGCCCAAGATGATCGCGCATCGCGCGC	531
Db	101	LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly	120
Qy	532	TTGATCAAGTGGCGCAGATGACACAGATCTCTGATCGAGATCTGCGCAAGAGGCCATC	591
Db	121	PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle	140
Qy	592	GGCACGTCGTGATCGGCGCCACCCCGTGAAACATCATCGCGCGCAACATGCTGACCC	651
Db	141	GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln	160
Qy	652	CTCGGCTGCAACCTGAATTCCTCCCATCAGCCCATCGAGACCGTGGCCGTGAAGCTGA	711
Db	161	IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys	180
Qy	712	CCCGCATGACCGGCGCCCAAGTGAAGCATGTGCCCTCGACCGAGAGAAAGATCAAGGCC	771
Db	181	ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla	200
Qy	772	CTGACCGCATCTCGGAGGATGGAGAGGGGCAAGATCACCAAGATCGCCCGCGAG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerIleGlyProGlu	220
Qy	832	AAACCTTACAAACCCCGTGTTCGCATCAAGAAGAAGACACCAAGTGGCGCAAG	891
Db	221	AsnProTyraAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys	240
Qy	892	CTGGTGAATTCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGCG	951
Db	241	LeuValAspPheArgGluLeuAsnArgTrpGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCACCCCGCGCTGAGAGAGAGAGAGCGTGCCTGTGGACGTGGGCGGCAC	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTACTTCAGCTGCCCTCGACGAGGACTTCGCAAGTACACCGCTTCACCATCCCC	1071
Db	281	AlaTyzPheSerValProLeuAspGluAspPheArgLysTyzThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACACGAGACCCCGGCATCCGCTACAGTACAACTGTGTGCCCCAGGGCTGG	1131
Db	301	SerIleAsnAsnGluThrProGlySerGlyTyzGlnTyzAsnValLeuProGlnGlyTrp	320
Qy	1132	AAGGGACGCCCCAGCATCTTCGAGCAGCATGATCCNAGATCTCTGGAGCCCTTCGCGCC	1191
Db	321	LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLys	340
Qy	1192	CGCAACCCCGAGATCGTGATCTACCACTATACAGCACTGTACGTGGCGCAGCACTG	1251
Db	341	GlnAsnProAspIleValIleTyzGlnTyzMetAspAspLeuTyzValGlySerAspLeu	360
Qy	1252	GAGATCGCGCAGCACCGCGCCAAAGATCGAGGAGCTGCGCAAGCACTGTGCTGGCTGGG	1311



```
QY 358 CAGGGCACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGC 411
Db      |||:::|
QY 412 ATCAAGTGGCGGCAGATCAAGGAGGCCCTGCTGGACACCGCGCCGACGACCGTG 471
Db      |||:::|
QY 472 CTGGAGGAGATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGGCATCGCGGC 531
Db      |||:::|
QY 532 TTCAATCAAGTGGCGGCAGTACGACGATCCTGATCGAGATCTCGCGCAAGAGGCCATC 591
Db      |||:::|
QY 592 GGCACCGTGTGATCGCGCCACCCCGTGAACATCATCGCGCGCGCAACATGCTGACCCAG 651
Db      |||:::|
QY 652 CTGGGCTGCACCTGAACCTTCCCATCAGCCCATCGAGACCTGCGCGTGAAGCTGAAG 711
Db      |||:::|
QY 712 CCGGCATGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAAGATCAAGGCC 771
Db      |||:::|
QY 772 CTGACCGCCATCTCGGAGGAGATGGAGAGAGGGCAAGATCACCAAGATCGCGCCCGAG 831
Db      |||:::|
QY 832 AACCCCTACACACCCCGCTTCGCCATCAAGAAGAGGACAGCACCAAGTGCSCGCAAG 891
Db      |||:::|
QY 892 CTGGTGAACCTTCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGC 951
Db      |||:::|
QY 952 ATCCCCCACCCGCGCCTGAGAGAGAGAGAGAGCGTGACCGTGTGGACGTGGCGGCAC 1011
Db      |||:::|
QY 1012 GCCTACTTTCAGCGTCCCTGGAGGAGACTTCGCGAAGTACACCGCTTCCACCATCCCC 1071
Db      |||:::|
QY 1072 AGCATCAACACGAGACCCCGGCATCCGCTACCAAGTACCAACGTGCTGCGCCCGCGGCTGG 1131
Db      |||:::|
QY 1132 AAGGGCAGCCCGACATCTTCCAGAGCAGATGATGACCAAGATCTCGAGAGCCCTTCCGCGCC 1191
Db      |||:::|
QY 1192 CGCAACCCCGAGATCGTGATCTACCACTGATCGACACCTGTACGTGGCGCAGCGACCTG 1251
Db      |||:::|
QY 1252 GAGATCGGCGCAGCACCCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGCTGCGCTGGGGC 1311
Db      |||:::|
QY 1312 TTCACACCCCGCAAGAGACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
Db      |||:::|
QY 1372 CTGACCCCGCAGCAGTGGCGCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGAGTGGACC 1431
Db      |||:::|
QY 1432 GTGAACGACATCCAGAGAGCTGGTGGCAAGCTGAACCTGGGCGCAGCCAGATCTACCCCGGC 1491
```

```
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyPrOgLy 440
QY 1492 ATCAAGTGGCGGCAGCTGTGCAAGCTGTGCGCGCGCAAGCCCTGACCGCATCGTG 1551
Db 441 IleLysValArgGlnLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1552 CCCTCACCAGGAGGCGCGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGCGAGCCC 1611
Db 461 ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1612 GTGCACGGGTGTACTACGACCCCAAGACCTGCTGGCCGAGATCCAGAGACGAGGC 1671
Db 481 ValHisGlyValTyTrpAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1672 CACGACAGTGGACCTTACCAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCGCAAG 1731
Db 501 GlnGlyGlnTrpThrTyrglnIleTyrglnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1732 TAGCCCAAGATGCGCACCCGCCACCAACGACGCTGAAGCAGCTGACCGAGGCGCGTGAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1792 AAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCCCTGCCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTyrglyLysThrProLysPheLysLeuProIle 560
QY 1852 CAGAAGGAGACCTGGGAGACCTGGTGACCGACTACTGGCAGCGCCACCTGGATCCCGGAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrrtrpGlnAlaThrTrpIleProGlu 580
QY 1912 TGGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTACAGCTGGAGAGAGGAGCCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTyrrtrpGlnLeuGluLysGluPro 600
QY 1972 ATCATCGGCGCCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrrValAspGlyAlaAlaAsnArgGluThrArgLeuGly 620
QY 2032 AAGCGCGGTACTGTGACCGACCGCGCGCGCAAGATCGTGAGCTGACCGAGACCCACC 2091
Db 621 LysAlaGlyTyrrLeuThrAsnLysGlyArgGlnLysValProLeuThrAsnThrThr 640
QY 2092 AACCAAGAAGACCGAGCTGCGAGCCATCCAGCTGGCCCTGCGAGACACGCGCAGCGAGTG 2151
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2152 AACATCGTGACCGACGACGATACGCCCTGGGGCATCTCCAGGCCCGCCCGACGAGAGC 2211
Db 661 AsnIleValThrAspSerGlnTyrrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSer 680
QY 2212 GAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTG 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuLysGlnLysValTyrrLeu 700
QY 2272 AGCTGGGTGCGCGCCCAAGGCGCATCGCGCGCAAGCAGAGCAGATCGACAGCTGGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2332 AAGGGATCGCAAGGTGTCTTCTGGACGGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

## RESULT 12

US-07-743-357-10

; Sequence 10, Application US/07743357

; Patent No. 5858646

; GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

ADDRESSEE: KIRBY EADES GALE BAKER  
 STREET: Box 3432, Station D  
 CITY: Ottawa  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: K1M 1H8  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/743,357  
 FILING DATE: 21-AUG-1991  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/CA90/00062  
 FILING DATE: 23-SEP-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gale, Edwin J.  
 REGISTRATION NUMBER: 28,584  
 REFERENCE/DOCKET NUMBER: 30924-2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (613) 237-6900  
 TELEFAX: (613) 237-0045  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1003 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: not relevant  
 TOPOLOGY: not relevant  
 MOLECULE TYPE: protein  
 HYPOTHEtical: NO  
 FRAGMENT TYPE: internal  
 ORIGINAL SOURCE:  
 ORGANISM: Human immunodeficiency virus type 1  
 STRAIN: ELI  
 US-07-743-357-10

Alignment Scores:  
 Pred. No.: 7,76e-211 Length: 1003  
 Score: 3497.50 Matches: 650  
 Percent Similarity: 95.83% Conservative: 40  
 Best Local Similarity: 90.28% Mismatches: 27  
 Query Match: 76.31% Indels: 3  
 DB: 2 Gaps: 2

US-09-610-313B-30 (1-2469) x US-07-743-357-10 (1-1003)

Qy 220 TTCTTCGGGAGGACCTGGCTTCCCGGAGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerProLysGln 20  
 Qy 280 AACCGGCCCAACGCCACCAGCGCGGAGCTGCGGCGC---GACAAACCCCGC 336  
 Db 21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40  
 Qy 337 AGCGAGGCGCGCGGAGCGCGGAGCCACCTG-----AACTTCCCGCAGATCACCCCTG 390  
 Db 41 SerLysThrGlyAlaGluArgGlnGlyThrValSerPheAsnProGlnIleThrLeu 60  
 Qy 391 TGGCAGCGCCCTTGGTGAGCATCAAGTGGCGGCGGCGGAGTCAAGAGGCGCTGCTGGAC 450  
 Db 61 TrpGlnArgProLeuValAlaIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80  
 Qy 451 ACCGGCCCGGACGACCGCTGCTGGAGAGATGAGCTGCCCGCAAGTGAGGCCCAAG 510  
 Db 81 ThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100  
 Qy 511 ATGATCGCGCGCTGCGGCTTCAATCAAGTGGCGGCGGCGGAGTCAAGAGGCGCTGCTGGAC 570  
 Db 101 MetIleGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleProIleGlu 120

Qy 571 ATCTGGCGCAAGAGGCCATCGGCACCGTGTCTGATCGGCCCGCCCGCGTGAACATCATC 630  
 Db 121 IleCysGlyGlnLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140  
 Qy 631 GGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACCTTCCCATCAGCCCATCAG 690  
 Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
 Qy 691 ACCGTGCCCTGGAAGCTGAAGCCCGCATCGAGCGGCGGCGGCGGCGGCGGCGGCGG 750  
 Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
 Qy 751 ACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGCGAGGAGATGGAGAGGAGGCGG 810  
 Db 181 ThrGluGlyLysIleLysAlaLeuThrGluIleCysThrAspMetGluLysGluGlyLys 200  
 Qy 811 ATCAACCAAGATCGGCCCGGAGAACCTCTACAACACCCCGCTGTTCCGCGCATCAAGAAG 870  
 Db 201 IleSerArgIleGlyProGluAsnProTyraAsnThrProIlePheAlaIleLysLys 220  
 Qy 871 GACAGCACCAAGTGGCGCAAGCTGTGGACTTCCCGAGCTGAAACAAGCGCACCCAGGAC 930  
 Db 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240  
 Qy 931 TTCTGGGAGGTGCGAGCTGGGCATCCCGCACCCCGCGCTGAGAGAGAGAGAGCGGTG 990  
 Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
 Qy 991 ACCGTGCTGACGTGGCGGCGGCTTCTTACGCGTCCCTGGAGGAGGACTTCCGCGAAG 1050  
 Db 261 ThrValLeuAspValGlyAspAlaTyraPheSerValProLeuAspGluAspPheArgLys 280  
 Qy 1051 TACACCGCTTACCATCCCGAGCATCAACAAGAGAGACCCCGCGCATCCGCTACCAGTAC 1110  
 Db 281 TyrThrAlaPheThrIleSerSerIleAsnAsnGluThrProGlyIleArgTyGlnTyr 300  
 Qy 1111 ACGTGTCTGCCCGGCGTGAAGGCGAGCCCGCGAGCATCTTCCAGAGCAGCATACCAAG 1170  
 Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 Qy 1171 ATCTGGAGGCTTCCGCGCGCGCAACCCCGAGATCTGTATCTACCATGATACATGAGCAG 1230  
 Db 321 IleLeuGluProPheArgLysGlnAsnProGluMetValIleTyrGlnTyrMetAspAsp 340  
 Qy 1231 CTGTACCTGGCGCAGCCTTGGAGATCGGCGAGCACCGCGCGCAAGATCGAGGAGCTGCGC 1290  
 Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLysLeuArg 360  
 Qy 1291 AAGCATCTGTCTGGCGCTTCCACCCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1350  
 Db 361 GluHisLeuLeuArgTrpGlyPheThrArgProAspLysLysHisGlnLysGluProPro 380  
 Qy 1351 TTCTGTGGATGGCTACGAGCTGCACCCCGAGAGTGGACCGCGCGAGCGCCATCGAGCTG 1410  
 Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400  
 Qy 1411 CCGGAGAGGAGAGCTGGACCGCTGAACGACATCCAGAAAGCTGTGGGCAAGCTGAACCTGG 1470  
 Db 401 ProGluLysGluSerTrpThrValAsnAspIleGlnAsnLeuValGluArgLeuAsnTrp 420  
 Qy 1471 GCAGCGCAGATCTACCCCGCGCATCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGG 1530  
 Db 421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLysLeuLeuArgGlyThr 440  
 Qy 1531 AAGGCGCTGACCGACATCGTGGCGCTGACCGAGGAGGCGGAGCTGGAGCTGGCGGAGAAC 1590  
 Db 441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460  
 Qy 1591 CGCGAGATCTTGGCGGAGCGGCTGCGACGGGTGTACTACGACCCCGAGCAAGAGCTGGT 1650  
 Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrTrpAspProSerLysAspLeuIle 480  
 Qy 1651 GCCGAGATCCAGAGAGGCGGCGGAGCGGAGCTGGACCTTACCATGATCTTACCGAGGCCCTTC 1710

Db 481 AlaGluileGlnLysGlnGlyHieGlyGlnTrpThrTyGlnIleTyGlnGluProPhe 500  
Qy 1711 AAGAACTGAAGACCGCAGTAGTACCCAGATGGCACCAGCCACCAACGAGCTGAAG 1770  
Db 501 LysAsnLeuLysThrGlyLysTyAlaArgMetArgGlyAlaHisThrAsnAspValLys 520  
Qy 1771 CAGCTGACCGAGGCGGTGACAGAGATCGCCATCGAGAGCATCGTGATCTGGGCAAGACC 1830  
Db 521 GlnLeuAlaGluAlaValGlnArgIleSerThrGluSerIleValIleTrpGlyArgThr 540  
Qy 1831 CCCAAGTTCGCGCTCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTACTGG 1890  
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpAlaGluTyTrp 560  
Qy 1891 CAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGG 1950  
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580  
Qy 1951 TACCAGCTGAGAGGAGGCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGCC 2010  
Db 581 TyGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyValAspGlyAlaAla 600  
Qy 2011 AACCGCGAGACCAAGATCGCAAGCGCGGTACGTGACCGACCGCGCGCGCAAGATC 2070  
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyValThrAspArgGlyArgGlnLysVal 620  
Qy 2071 GTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGGTGAACACCCCGCCCTGGTGAAGCTGTGG 2130  
Db 621 ValProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640  
Qy 2131 CAGGACAGCGCGAGGAGTGAACATCGTACCGACAGCAGCGAGTACCGCGCGCGCGCATCATC 2190  
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIleIle 660  
Qy 2191 CAGCGCCAGCGCGCAAGAGCGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATC 2250  
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIle 680  
Qy 2251 AAGAGAGAGAGTGTACCTGAGCTGGTGGTGGCGCGCGCGCGCGCGCGCGCGCGAG 2310  
Db 681 LysLysGlnLysValTyLeuAlaTrpValProAlaHisLysGlyIleGlyAsnGlu 700  
Qy 2311 CAGATCGACAGCTGTGTGAGCAGGCGATCCGCAAGGTGTGTTCCTGGACGCGCATCGAT 2370  
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 13  
US-07-743-357-8  
; Sequence 8, Application US/07743357  
; Patent No. 5858646  
; GENERAL INFORMATION:  
; APPLICANT: Kang, Yong C.  
; TITLE OF INVENTION: Polypeptide having immunological  
; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: KIRBY EADES GALE BAKER  
; STREET: Box 3432, Station D  
; CITY: Ottawa  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: K1M 1H8  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07743,357  
; FILING DATE: 21-AUG-1991  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA90/00062  
; FILING DATE: 23-FEB-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gale, Edwin J.  
; REGISTRATION NUMBER: 28,584  
; REFERENCE/DOCKET NUMBER: 30924-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (613) 237-6900  
; TELEFAX: (613) 237-0045  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1003 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; ORGANISM: Human immunodeficiency virus type 1  
; STRAIN: RF  
; US-07-743-357-8  
  
Alignment Scores:  
Pred. No.: 1,2e-210 Length: 1003  
Score: 3494.50 Matches: 651  
Percent Similarity: 95.83% Conservative: 39  
Best Local Similarity: 90.42% Mismatches: 27  
Query Match: 76.25% Indels: 3  
DB: 2 Gaps: 2  
  
US-09-610-313B-30 (1-2469) x US-07-743-357-8 (1-1003)  
Qy 220 TTCTTCGGGAGGACCTGGCTTCCCGAGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluLeuSerSerGluGln 20  
Qy 280 AACCGCGCGCAACAGCCCGACCGCGAGCTCGAGTGGCGGC---GACAAACCCCGCGC 336  
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSerLeu 40  
Qy 337 AGCGAGCGCGCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390  
Db 41 SerGluAlaGlyLysArgGlnGlyThrValSerPheSerPheProGlnIleThrLeu 60  
Qy 391 TGCAGCG 450  
Db 61 TrpGlnArgProIleValThrValLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80  
Qy 451 ACCGCGCGCGCGAGCACCGCTGCTGGAGGAGATGAGCTGCCCGCGCAAGTGGAGCGCGCAAG 510  
Db 81 ThrGlyAlaAspAspThrValLeuGluMetAsnLeuProGlyLysTrpLysProLys 100  
Qy 511 ATGATCGCGCGCGATCGCGCGCTTCATCAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 570  
Db 101 MetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu 120  
Qy 571 ATCTGCGCGCAAGAGCGCATCGCGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630  
Db 121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140  
Qy 631 GCGCGCAACATGCTGACCCCGAGTGGGCTGCACCTCGACCTTCCCATCAGCCCCCATCGAG 690  
Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
Qy 691 ACCGTGCGCGTGAAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750  
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
Qy 751 ACCGAGGAGAGAGATCAAGCG 810  
Db 181 ThrGluGlnLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLys 200

QY 811 ATCCAGATCGCCGCGGAGAACCCCTACACACCCCGGTGCGCATCAAGAGAG 870  
Db : : : : :  
QY 201 IleserlylleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
Db : : : : :  
QY 871 GACAGCAACAAAGTGGCGAAGCTGTGACTTCGCGAGCTCAACAAAGCGCACCCAGGAC 930  
Db : : : : :  
QY 221 AspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240  
QY 931 TTCTGGAGGTGAGCTGGGATCCCCACCCCGCGCGCTGAGAGAAAGAGCGGTG 990  
Db : : : : :  
QY 241 PheTrpGluValGlnLeuGlyLysProHisProAlaGlyLeuLysLysSerVal 260  
QY 991 ACCGTGCTGGACGTGGGCGAGCTTACTTCAAGCTGCCCTGGACGAGGACTTCGCAAG 1050  
Db : : : : :  
QY 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysGluPheArgLys 280  
QY 1051 TACACCGCTTCAACCATCCCGAGATCAACACAGACACCCCGCATCCGCTACCAAGTAC 1110  
Db : : : : :  
QY 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProArgLysGlyThrGlnTyr 300  
QY 1111 AACGTGCTGCCCGAGGCTGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATCAACAAAG 1170  
Db : : : : :  
QY 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
QY 1171 ATCTGGAGCCCTTCCGCGCGGCAACCCCGAGATCGTGATCTACCATGATGAGCAG 1230  
Db : : : : :  
QY 321 IleLeuGluProPheLysLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1231 CTGTACGTGGCGAGCATCGGCGAGCGCAGCAGCGCGCAAGATCGAGGAGCTGGCG 1290  
Db : : : : :  
QY 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgIleLysLysLysGluLeuArg 360  
QY 1291 AAGCACCTGTGCTGGGCTTCCACCAACCCCGAGAGAGAGAGAGAGAGAGAGAGAG 1350  
Db : : : : :  
QY 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
QY 1351 TTCTGTGGATGGCTAGAGCTGACCCCGAGAGTGGACCGTGGAGCCCATCGAGCTG 1410  
Db : : : : :  
QY 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeu 400  
QY 1411 CCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGGCGCAAGCTCAACTGG 1470  
Db : : : : :  
QY 401 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
QY 1471 GCCAGCAGATCTACCCCGCATCAAGGTGGCCAGCTGTGCAAGCTGCTGCGCGGCGCC 1530  
Db : : : : :  
QY 421 AlaSerGlnIleTyrAlaGlyLysValLysGlnLeuCysLysLeuLeuArgGlyThr 440  
QY 1531 AAGCCCTGACGACATCGTCCCTGACCGAGAGCGCGAGCTGGAGCTGGCGGAGAAC 1590  
Db : : : : :  
QY 441 LysAlaLeuThrGluValValGlnLeuThrLysGluAlaGluLeuLeuAlaGluAsn 460  
QY 1591 CGCAGATCTCGCGGAGCGCTGACGGCGGTGTACTACGACCCCGAGGAGCTGGTG 1650  
Db : : : : :  
QY 461 ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
QY 1651 GCCGAGATCCAGAGCAGGCGCACCGCAGTGGACCTACAGATCTACGAGGAGCCCTTC 1710  
Db : : : : :  
QY 481 AlaGluIleGlnLysGlnGlyGlnTrpThrThrGlnIleTyrGlnLysProPhe 500  
QY 1711 AAGAACTGAAGCCGCAAGTACGCCAAGATGCCACCGCCACCAACAGCAGCTGAAG 1770  
Db : : : : :  
QY 501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520  
QY 1771 CAGCTGACCGAGCGCTGCAGAAATGCCATGAGAGCATCGTGATCTGGGGGAGAGACC 1830  
Db : : : : :  
QY 521 GlnLeuThrGluAlaValGlnLysValAlaThrGluSerIleValIleTrpGlyLysThr 540  
QY 1831 CCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGCTGGACCGACTACTGG 1890  
Db : : : : :  
QY 541 ProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 560  
QY 1891 CAGGCCACCTGGATCCCGCGAGTGGAGTTCTGTGAACACCCCGCCCTCGTGGTGAAGCTGG 1950

Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
QY 1951 TACCAGCTGGAGAGAGAGCCCATCATCGCGCCCGAGACCTTCTAGCTGGAGCGGCCGCC 2010  
Db : : : : :  
QY 581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 2011 AACCCGAGAGCAAGATCGGCAAGCGCGCTACGTACCGACCGCGGCGCGAGAGATC 2070  
Db : : : : :  
QY 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
QY 2071 GTGAGCTGACCGAGAGACCAACAGAGACCGAGCTGACGAGCATCCAGCTGGCCCTG 2130  
Db : : : : :  
QY 621 ValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2131 CAGGACGCGCAGCAGGAGTGAACATCGTGACCGACAGCAGCAGCTCGGCGCATCATC 2190  
Db : : : : :  
QY 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
QY 2191 CAGCCCGAGCGCGAGAGGAGGAGGAGCTGTGTGAACAGATCATCGAGCAGCTGATC 2250  
Db : : : : :  
QY 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680  
QY 2251 AAGAAGAGAGAGGTGTACTGAGCTGGGTGCGCCCGCAGAGGAGCATCGCGGCAACGAG 2310  
Db : : : : :  
QY 681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysGlnGlu 700  
QY 2311 CAGATCCACAGAGTGTGTGAGCAGGAGCATCGCAGAGGTGTCTTCTGGAGCGGATCGAT 2370  
Db : : : : :  
QY 701 GlnValAspArgLeuValSerThrGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

## RESULT 14

US-07-743-357-7

; Sequence 7, Application US/07743357

; Patent No. 5858646

; GENERAL INFORMATION:

; APPLICANT: Kang, Yong C.

; TITLE OF INVENTION: Polypeptide having immunological

; TITLE OF INVENTION: activity for use as diagnostic reagent and/or vaccine

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: KIRBY EADES GALE BAKER

; STREET: Box 3432, Station D

; CITY: Ottawa

; STATE: Ontario

; COUNTRY: Canada

; ZIP: K1M 1H8

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA: US/07743, 357

; APPLICATION NUMBER: US/07743, 357

; FILING DATE: 21-AUG-1991

; CLASSIFICATION: 424

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PCT/CA90/00062

; FILING DATE: 23-FEB-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Gale, Edwin J.

; REGISTRATION NUMBER: 28,584

; REFERENCE/DOCKET NUMBER: 30924-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (613) 237-6900

; TELEFAX: (613) 237-0045

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1004 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein



; HYPOTHETICAL: NO  
 ; FRAGMENT TYPE: internal  
 ; ORIGINAL SOURCE: Human immunodeficiency virus type 1  
 ; ORGANISM: SF2  
 ; STRAIN: US-07-743-357-7

Alignment Scores:

Pred. No.: 3,52e-210 Length: 1004  
 Score: 3487.00 Matches: 650  
 Percent Similarity: 95.42% Conservative: 38  
 Best Local Similarity: 90.15% Mismatches: 29  
 Query Match: 76.09% Indels: 4  
 DB: 2 Gaps: 2

US-09-610-313B-30 (1-2469) x US-07-743-357-7 (1-1004)

QY	220	TTCTTCGCGAGGACCTGGCTTCCCTCCCGAGGCAAGCCCGAGGATTCCTCCAGCGAGCAG	279
DB	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
QY	280	AACGGGGCGCAACAGCCACACAGCGCGGAGCTGAGTGGCGGC-----GACACCC	333
DB	21	ThrArgAlaAsnSerProThrArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer	40
QY	334	CGCAGCGAGCGCGCGCGAGCGCCAGCGCACCTCG-----AACTTCCCGCCAGATCAC	387
DB	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr	60
QY	388	CTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGGGCGGCAGATCAAGAGGCGCTGCTG	447
DB	61	LeuTrpGlnArgProLeuValThrIleArgIleGlyGlnLeuLysGluAlaLeuLeu	80
QY	448	GACACCGCGCGCAGCAGACCTGCTGGAGAGATGAGCTGCGCGGCAAGTGGAGCC	507
DB	81	AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro	100
QY	508	AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCAGTACGACACAGATCCTGATC	567
DB	101	LysMetIleGlyIleGlyIleGlyPheIleLysValArgGlnIleProVal	120
QY	568	GAGATCTGCGCGCAAGAGCGCATCGGCACCTGTGATCGCGCCCGCCACCCCGTGAACATC	627
DB	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
QY	628	ATCGCGCGCAACATGCTGACCCAGCTGGCTGGCTGACCTGAACTTCCCATCAGCCCATC	687
DB	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
QY	688	GAGACCGTGCCTGAGCTGAAGCGCGCATGAGCGCGCCCAAGGTGAAGCAGTGGSCC	747
DB	161	GluThrValProValLysLeuLysProGlyWecAspGlyProLysValLysGlnTrpPro	180
QY	748	CTGACCGAGGAGAGATCAAGCGCCCTCAGCGCCATCTGCGAGGAGATGGAAGAGGGC	807
DB	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
QY	808	AAGATCACCAAGATCGCGCGCGAGAACCCCTACAAACCCCGCTGTTCCGCTCATCAAGA	867
DB	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220
QY	868	AAGGACAGCACCAAGTGGCGCAGCTGGTGGACTTCCGCGAGCTGGAACAGCGCCAC	927
DB	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
QY	928	GACTTCTGGGAGTGCAGCTGGGCATCCCCACCCCGCGCTGGAAGAGAGAGAGAGC	987
DB	241	AspPheTrpGluValGlnLeuGlyIleProIleProAlaGlyLeuLysLysLysSer	260
QY	988	GTGACCGTGTGACGTGGGCGAGCGCTTCTAGCGTGGCTGGAGAGGACTTCCGC	1047
DB	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg	280

QY	1048	AAGTACACCGCCTTACCATCCCGAGCATCAACACAGACGCCCGGCTCCGCTACCAG	1107
DB	281	LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln	300
QY	1108	TACAACGTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC	1167
DB	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
QY	1168	AAGATCCTCGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACAGATACATGGAC	1227
DB	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	340
QY	1228	GACCTGTAGTGGCGCAGCAGCTGGAGATCGGCGCACCGCGCCCAAGATCGAGGAGCTG	1287
DB	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu	360
QY	1288	CGCAACGACCTGCTGGCTGGGCTTACCAACCCCGAGAGAGACACAGAGAGGCC	1347
DB	361	ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
QY	1348	CCCTTCTGTGGATGGCTACGAGCTGCACCCCGACCAAGTGGACCTGCGAGCCCATCGAG	1407
DB	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet	400
QY	1408	CTGCCCGAGAGGAGAGCTGACCCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAAC	1467
DB	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
QY	1468	TGGGCGAGCGAGATCTACCCCGGCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGGC	1527
DB	421	TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly	440
QY	1528	GCCAAAGCCTGACCGCACATCGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAG	1587
DB	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu	460
QY	1588	AACCGGAGATCTCGCGGAGCCCGTGCACGCGCTGTACTACACCCCGAGAGGAGCTG	1647
DB	461	AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu	480
QY	1648	GTGGCGAGATCCAGAGCGGCGCAGCAGTGCACCTACCGATCTACCGAGGAGCCC	1707
DB	481	ValAlaGluIleGlnLysGlnGlyGlnIleTrpThrTyrGlnIleTyrGlnGluPro	500
QY	1708	TTCACGAACCTGAAGACCGCAAGTACGCCAAGATCGCACCGCCCGCACCAACAGCAGTG	1767
DB	501	PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal	520
QY	1768	AAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATCGAGAGCATCGTGTCTGGGGCAAG	1827
DB	521	LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys	540
QY	1828	ACCCCGAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTAC	1887
DB	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpMetGluTyr	560
QY	1888	TGGCGCGCACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG	1947
DB	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	580
QY	1948	TGCTACAGCTGGAAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCC	2007
DB	581	TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	600
QY	2008	GCCAAACCGGAGACCAAGATCGGCAAGCGCGCTAGTGAACCGAGCCGCGGCGGAGAG	2067
DB	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys	620
QY	2068	ATCGTGCAGCTGACCGAGACCCCAACAGACCGAGCTGCAGGCGCATCTTCCAGCTGGCC	2127
DB	621	ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2128	CTGCAGGACAGCGCGCAGGAGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC	2187



```

Db      441  LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu  460
QY      1567  GCCGAGCTGAGCTGGCGGAGAACCCGAGATCTCGCGGAGCCCGTGCACGGCGTGTAC  1626
Db      461  AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr  480
QY      1627  TAGCACCCACAGGACCTGGTGGCGGAGATCCAGAAGCAGGGCCACACGACCATGGACC  1686
Db      481  TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnTyrThr  500
QY      1687  TACAGATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGGC  1746
Db      501  TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg  520
QY      1747  ACCGCCACACACGACGAGCTGACGAGCTGACGAGCGCCGTGCAGAGATCGCATCGAG  1806
Db      521  SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu  540
QY      1807  AGCATCGTGATCTGGGCAAGACCCCAAGTTCGCGTCCGCTCCATCCAGAAGGAGACCTGG  1866
Db      541  SerIleValIleTyrGlyLysThrProLysPheArgLeuProValGlnLysGluValTyr  560
QY      1867  GAGACCTGTGTGGACCGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAAC  1926
Db      561  GluAlaTyrTyrThrAspHisTyrGlnAlaThrTyrIleProGluTyrGluPheValAsn  580
QY      1927  ACCCCCCCTGTGTGAGCTGTGTACGACCTGAGAGGAGAGGCCCATCATCGGGGCGGAG  1986
Db      581  ThrProProLeuValLysLeuTyrTyrGlnLeuGluThrGluProIleSerGlyAlaGlu  600
QY      1987  ACCTTCTACGTGGACGGCGGCCCAACCGGAGACCAAGATCGCAAGCGCGGTACGTG  2046
Db      601  ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal  620
QY      2047  ACCGACCGGGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAG  2106
Db      621  ThrAspArgGlyArgGlnLysValIleSerIleAlaAspThrThrAsnGlnLysAlaGlu  640
QY      2107  CTGAGGCCATCCAGCTGGCGCTGCAGGACGCGGCGAGCGGTGACATCGTGACCGAC  2166
Db      641  LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp  660
QY      2167  AGCCAGTACGCCCTGGGCATCATCCAGGCCGAGCCGACAGAGCGAGCGAGCTGGTG  2226
Db      661  SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal  680
QY      2227  AACGAGATCATCGAGAGCTGATCAAGAGGAGAGGTGTACTGTAGCTGGTGCCCGCC  2286
Db      681  SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTrpValProAla  700
QY      2287  CACAAGGGCATCGCGGCAAGCAGCATCGACAAGCTGGTGAGCAAGGGCATCCGCAAG  2346
Db      701  HisLysGlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys  720
QY      2347  GTGCTGTCTCTGGACGGCATCGAT  2370
Db      721  IleLeuPheLeuAspGlyIleGlu  728

```

Search completed: June 2, 2005, 05:11:22  
Job time : 112.083 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 04:23:54 ; Search time 153.54 Seconds

(without alignments)  
11117.427 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 4583

Sequence: 1 gtcagccaccatgcccga.....gggctagcaccggtgaattc 2469

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters: 2931222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO spooll/US09610313/runat\_31052005\_15138\_15193/app\_query.fasta\_1.7893  
-DB=Published Applications AA -QFMT=fastan SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCU=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09610313 @CGN 1 1 550 @runat 31052005\_15138\_15193  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pcp.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pcp.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pcp.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp.\*
  - 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
  - 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\* Query  
Result

No.	Score	Match	Length	DB	ID	Description
1	3608	78.7	998	16	US-10-332-413-4	Sequence 4, Appli
2	3539.5	77.2	850	16	US-10-332-413-10	Sequence 10, Appl
3	3535	77.1	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3513	76.7	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3485	76.0	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3485	76.0	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3466	75.7	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3466	75.6	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3465	75.6	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3465	75.6	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3457	75.4	995	15	US-10-296-734-2	Sequence 2, Appli
12	3443	75.1	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3289.5	71.8	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3221.5	70.3	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3135.5	68.4	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3135.5	68.4	1350	17	US-10-380-641-35	Sequence 35, Appl
17	3135.5	68.4	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2860.5	62.4	875	10	US-09-952-060-6	Sequence 6, Appli
19	2860.5	62.4	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2860.5	62.4	875	16	US-10-380-641-6	Sequence 6, Appli
21	2860.5	62.4	875	17	US-10-636-730-6	Sequence 6, Appli
22	2855	62.3	850	10	US-09-952-060-2	Sequence 2, Appli
23	2855	62.3	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2855	62.3	850	16	US-10-380-641-2	Sequence 2, Appli
25	2855	62.3	850	17	US-10-636-730-2	Sequence 2, Appli
26	2827	61.7	562	9	US-09-735-487-14	Sequence 14, Appl
27	2827	61.7	562	14	US-10-342-188-14	Sequence 14, Appl
28	2823	61.6	560	15	US-10-399-920-2	Sequence 2, Appli
29	2821	61.6	561	14	US-10-059-271-87	Sequence 87, Appl
30	2815	61.4	562	14	US-10-102-622-10	Sequence 10, Appl
31	2814.5	61.4	875	10	US-09-952-060-8	Sequence 8, Appli
32	2814.5	61.4	875	15	US-10-168-217A-8	Sequence 8, Appli
33	2814.5	61.4	875	17	US-10-380-641-8	Sequence 8, Appli
34	2814.5	61.4	875	17	US-10-636-730-8	Sequence 8, Appli
35	2809	61.3	850	10	US-09-952-060-4	Sequence 4, Appli
36	2809	61.3	850	15	US-10-168-217A-4	Sequence 4, Appli
37	2809	61.3	850	16	US-10-380-641-4	Sequence 4, Appli
38	2809	61.3	850	17	US-10-636-730-4	Sequence 4, Appli
39	2796	61.0	560	14	US-10-205-641-1	Sequence 1, Appli
40	2786.5	60.8	561	14	US-10-102-622-12	Sequence 12, Appl
41	2771	60.5	560	11	US-09-725-652-1	Sequence 1, Appli
42	2740	59.8	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2727	59.5	979	14	US-10-271-181B-118	Sequence 118, Appl
44	2626	57.3	851	14	US-10-059-271-88	Sequence 88, Appl
45	2609	56.9	999	15	US-10-346-000A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-10-332-413-4  
; Sequence 4, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; TITLE OF INVENTION: Subtype Isolates, Their Genes nd Modifications and Derivatives T  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924

; PRIOR FILING DATE: 2000-09-15									
; NUMBER OF SEQ ID NOS: 32									
; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO 4									
; LENGTH: 998									
; TYPE: PRT									
; ORGANISM: Human immunodeficiency virus type 1									
US-10-332-413-4									
Alignment Scores:									
Pred. No.: 3.5e-168 Length: 998									
Score: 3608.00 Matches: 670									
Percent Similarity: 97.63% Conservative: 29									
Best Local Similarity: 93.58% Mismatches: 17									
Query Match: 78.73% Indels: 0									
DB: 16 Gaps: 0									
US-09-610-313B-30 (1-2469) x US-10-332-413-4 (1-998)									
QY	223	TTCCGGGAGGACCTGGCTTCCCGCAGGGCAAGGCCCGCGAGTTCCTCCAGCAGCAGAAC	282	Db	261	ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe	280		
Db	1	PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr	20	QY	1063	ACCATCCCGAGCATCAACACAGAGAGCCCGCGCATCGGTACACAGTACAGCTGCTGCC	1122		
QY	283	CGCGCCAAACAGCCCAACAGCCGAGCTGCAGTGGCGCGCGCAACACCCCGCAGCGAG	342	Db	281	ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro	300		
Db	21	ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu	40	QY	1123	CAGGGCTGGAAGGGCAGCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTCTGGAGCCC	1182		
QY	343	GCGGCGCGCAGCGCCAGCGGACCCCTGAATCTCCCGCAGATCACCTGTGCGAGCGCCC	402	Db	301	GlnGlyTyrPheGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro	320		
Db	41	ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTyrGlnArgPro	60	QY	1183	TTCCGGCGCCGCAACCCCGAGATCGTGATCTACCATGATGACGACCTGTAGTGGGC	1242		
QY	403	CTGGTGAGCATCAAGTGGCGGCGCAGATCAAGAGGCCCTGCTGGACACCGCGCGCGAC	462	Db	321	PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly	340		
Db	61	LeuValSerIleLysIleGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp	80	QY	1243	AGCGACTGAGATCGCGCAGCACCAGCGCAAGATCGAGAGTGGCAGACCTGCTG	1302		
QY	463	GACACCGTGTGGAGGAGATGAGCTGCCCGCAAGTGAAGCCCAAGATGATCGGCGGC	522	Db	341	SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHisLeuLeu	360		
Db	81	AspThrValLeuGluAspIleAsnLeuProGlyLysTyrLysProLysMetIleGlyGly	100	QY	1303	CGCTGGGCTTCCACACCCCGACCAAGAGCACCAGAGGAGCCCGCTTCTGTGTGATG	1362		
QY	523	ATCGCGGCTTCATCAAGTGGCGCAGTACGACGAGATCTGATCGAGATCTCGGCAAG	582	Db	361	LysTyrPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTyrMet	380		
Db	101	IleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLys	120	QY	1363	GGCTACGAGCTGCACCCGACAAAGTGGACCGTCCAGCCCATCGAGTGGCCGAGAGAG	1422		
QY	583	AAGGCCATCGGACCGTGTGATCGGCGCCCAACCCCGTGAACATCATCGGCGCCGCAATG	642	Db	381	GlyTyrGluLeuHisProAspLysTyrThrValGlnProIleGlnLeuProGluLysAsp	400		
Db	121	LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet	140	QY	1423	AGCTGACCGTGAACCATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCAGCCAGATC	1482		
QY	643	CTGACCCAGCTGGGCTGCACCTGAATCTCCCATCAGCCCATCGAGACCGTGGCGGTG	702	Db	401	SerTyrThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTyrAlaSerGlnIle	420		
Db	141	LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal	160	QY	421	TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr	440		
QY	703	AAGTGAAGCCCGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAG	762	QY	1543	GACATCGTGCCTGACCGAGGAGGCGAGCTGAGAGTGGCGCGAGAACCCGAGATCCTG	1602		
Db	161	LysLeuLysProGlyMetAspGlyProLysValLysGlnTyrProLeuThrGluLys	180	Db	441	AspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluLeuLeu	460		
QY	763	ATCAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAGGAGGCGCAGATCACCAAGATC	822	QY	1603	CGCGAGCCGTGACCGGGTGTACTACGACCCCGCAGCAGGACCTGGTGGCGAGATCCAG	1662		
Db	181	IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle	200	Db	461	LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln	480		
QY	823	GGCCCCGAGAACCCCTACAACACCCCGCTGTGCCCATCAAGAGAGGAGGACACCAAG	882	QY	1663	AAGCAGGCGCACGACAGTGGACCTACCATGATCTACGAGAGCCCTTCAAGAACCTGAAG	1722		
Db	201	GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys	220	Db	481	LysGlnGlyAspAspGlnTyrThrTyrGlnIleTyrGlnGluProPheLysLeuLys	500		
QY	883	TGGCGCAAGCTGGTGAATCTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTG	942	QY	1723	ACCGCAAGTACGCCAAGATGCGCAGCGCCCAACCAACGACGTGAGAGCTGACCGAG	1782		
Db	221	TyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTyrGluVal	240	Db	501	ThrGlyLysTyrAlaLysArgArgThrThrHisThrAsnAspValLysGlnLeuThrGlu	520		
QY	943	CAGCTGGGCATCCCCACCGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1002	QY	1783	GCGTGCAGAGATCGCCATGGAGAGCATCGTGTACTTGGGGCAGACCCCGAGTTCGCG	1842		
Db	241	GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAsp	260	Db	521	AlaValGlnLysIleSerLeuGluSerIleValIleTyrGlyLysThrProLysPheArg	540		
QY	1003	GTGGGCGACGCTACTTTCAGCTGCGCCCTGAGCAGGAGCTTCCGCAAGTACACCGCTTC	1062	QY	1843	CTGCCCATCCAGAGGAGACCTCGGAGACCTGTGTGGACCGCTACTTGGCAGGCGCACCTGG	1902		



```
QY 1360 ATGGGCTACGAGTGCACCCCGCAAGTGGACCGTCAGCCCATCGAGCTGCCCGGAGAAG 1419
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
520 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLys 539
QY 1420 GAGAGCTGGACCGTGAACACGATCCAGAACCTGGTGGCGAAGCTGAATCGGCGACCCAG 1479
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
540 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpThrSerGln 559
QY 1480 ATCTACCCCGGCTCAAGGTGCGCCAGCTGTGAAGCTGCTGCGCGCGCCAGGCCCTG 1539
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 579
QY 1540 ACCGACATCTGCCCTCGACCGGAGGCGCCAGCTGGAGCTGCGCCGAGAACCCGAGATC 1599
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIle 599
QY 1600 CTGCGGAGCCCGTGCACCGCGTGTACTACGACCCCGCAAGCACTGTGTGGCGGAGATC 1659
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
600 LeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIle 619
QY 1660 CAGAAGCAGGCGCCACGACGAGTGGACCTACGATCTACGAGGACCCCTTCAAGAACCTG 1719
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
620 GlnLysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 639
QY 1720 AAGACCGGCAAGTACGCAAGATGCGCACCGCCGACACCAACGACGTGAAGCACTGACC 1779
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
640 LysThrGlyLysTyrAlaLysArgArgThrHisThrAsnAspValLysGlnLeuThr 659
QY 1780 GAGGCGGTGCAGAAAGTTCGCCATCGGAGAGCATCGTATCTGGGGCGAAGACCCCAAGTTC 1839
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 GluAlaValGlnLysIleSerLeuGluSerIleValThrTrpGlyLysThrProLysPhe 679
QY 1840 CGCTGCCCATCAGAAAGGAGACTGGGAGACCTGTGGACCCACTGTGGCGGCGCCACC 1899
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 ArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpThrAspTyrTrpGlnAlaThr 699
QY 1900 TGGATCCCGGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGGTACCAGCTG 1959
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
700 TrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 719
QY 1960 GAGAAGGAGCCCATCATCGCGCGGAGACCTTCTAGTGTGACCGCGCGCCGCAACCGCGAG 2019
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
720 GluLysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 739
QY 2020 ACCAAGTCCGCAAGCGCGCTAGTGCACCGCGCGCGCGCGAGATCTGTGAGCCCTG 2079
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
740 ThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeu 759
QY 2080 ACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCGCTGCAGGACAGC 2139
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 SerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 779
QY 2140 GGCAGCGAGTGAACATCGTGACCGACGACGAGTACGCCCTGGCGCATCATCCAGGCCAG 2199
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
780 GluSerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 799
QY 2200 CCGGACAAGACGAGCAGCGCTGCTGAACAGATCATCAGACGAGCTGATCAAGAGGAG 2259
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
800 ProAspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuLysLysGlu 819
QY 2260 AAGGTGTACCTGAGCTGGGTGCCCGCCACAAAGGGCATCGCGCGCAACGAGAGATCGAC 2319
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
820 ArgAlaTyrLeuSerTrpValProAlaHisGlyIleGlyLysGluGlnValAsp 839
QY 2320 AAGTGTGTGACGAGGCGCATCCGCAAGGTGCTG 2352
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
840 LysLeuValSerSerGlyIleArgLysValLeu 850
```

## RESULT 3

```
US-10-283-847-17
; Sequence 17, Application US/10283847
; Publication No. US20030162720A1
; GENERAL INFORMATION:
221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
```

```
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Walker, Bruce
; TITLE OF INVENTION: THERAPEUTIC ANTI-HIV (IV9) COMPOUNDS
; FILE REFERENCE: GZ 2112.00
; CURRENT APPLICATION NUMBER: US/10/283,847
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 60/345,116
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
US-10-283-847-17
```

```
Alignment Scores:
Pred. No.: 1,28e-164 Length: 1003
Score: 3535.00 Matches: 659
Percent Similarity: 96.12% Conservative: 34
Best Local Similarity: 91.40% Mismatches: 24
Query Match: 77.13% Indels: 4
DB: 14 Gaps: 2
```

US-09-610-313B-30 (1-2469) x US-10-283-847-17 (1-1003)

```
QY 220 TTCTTCGCGGAGACCTGCGCTTCCCCAGGGCAAGSCCGCGAGTTCCCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCAAACACGCCCCACCGCGCGAGCTCGAGGTG-----CGCGCGCAACACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGAGCCCGAGCGGACCCCTG-----AACTTCCCCAGATCACC 387
Db 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
QY 388 CTGTGCGCGCGCCCTGTTGAGCATCAAGTGGCGGCGCAGATCAAGGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCACACCCGTGTGTGGAGGAGATGAGCTGCCCGCAAGTGGAGACCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluWetSerLeuProGlyArgTrpLysPro 100
QY 508 AGATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGGCGAGTACGACGAGATCCTGATC 567
Db 101 LysMetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
QY 568 GAGATCTGCGGCAAGAGCCATCGGCACCGCTGATCGGCCCGCCACCCCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGGTGCACCTGACCTGAACCTTCCCATCAGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGAGTGAAGCGCGGATCGAGCGGCCCGCAAGGTGAAGCAGTGGCCCC 747
Db 161 GluThrValProValLysLeuLysProGlyWetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGGAGAGATCAAGGCCCTGACCGCCATCTCGCGAGAGATGGAAGAGAGGGGC 807
Db 181 LeuThrGluGlnLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCCCGCGAGAACCCTTACAACACCCCGCTTTCGCCCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGAGCTTCCGCGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
```



```
QY 928 GACTTCTGGAGGTGCTAGCTGGGATCCCTCCCAACCCCGCCGCTTGAAGAAAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGGAGCGGCGGAGCTTCTACGCTGCGGCTTGGAGGAGCTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 280
QY 1048 AAGTACACCGCTTTCACCATCCAGCATCAACACGAGACCCCGGATCCGCTACACAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnGlnThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGCTGCCCGGAGGTGAGGGGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerMetThr 320
QY 1168 AAGATCTCTGGAGCCTTCCCGCGCCGCAACCCGAGATCGTGATCTACAGTACATGAC 1227
Db 321 LysIleLeuGluProPheArgGlyGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1228 GACCTGTACGTGGGAGGAGCTCGAGATCGGCGCAGCACCCGCGCAAGATCGAGGAGCTG 1287
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 360
QY 1288 CGCAAGCACCTGCTGCGCTGGGCTTACCAACCCCGCAAGAGACCAAGAGGAGCC 1347
Db 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCCTTCTGTGGTGGCTAGAGCTGCACCCGCAAGTGCAGCGGCGGAGCCCATCGAG 1407
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1408 CTGCCCCGAGAGGAGCTGACCGCTGAACACATCCAGAGCTGGTGGGCAAGCTGAAC 1467
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1468 TGGCCGACCGAGATCTACCCCGGATCAAGTGGCCGAGCTGTCAAGCTGCTCGCGGC 1527
Db 421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1528 GCCAAGCCCTGACGACATGCTCCCTGACCGGAGGCGGAGCTGGAGCTGGCGGAG 1587
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
QY 1588 AACCGGAGATCTCTGCGGAGCCGCTGCACCGGCTGTACTACGACCCCGAGCAAGACTG 1647
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1648 GTGCCCGGAGATCCAGAGCAGGCGCCACGACCATCGAGTGGACCTACCATGATCTTACCAGGAGCC 1707
Db 481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
QY 1708 TTCAGAACCTGAGACCGGCAAGTACGCCAAGATGGCCACCGCCGCAACCAAGAGCTG 1767
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGGCGCTGAGAGATGCCATGCCATGGAGAGCATCGTGATCTGGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540
QY 1828 ACCCCCAAGTTCGCTGCCCATCCAGAAAGAGACCTGGGAGACCTGGTGACCGACTAC 1887
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr 560
QY 1888 TGGCAGGCGCCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1947
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1948 TGGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGGCGC 2007
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
```

```
QY 2008 GCCAACCGCAGAGACCAAGATCGGCAAGCCCGGCTACGTGACCGACCGCGGCGGCAAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2068 ATCTGTAGCTTCCGAGAGACCAACCAAGAGACCGAGCTGCGAGCCATCCAGCTGGCC 2127
Db 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
QY 2128 CTGACGACGACGCGCAGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCGCCGCGCAGAGAGGAGAGCGAGCTGTGTGAACCAAGATCATCGAGCTG 2247
Db 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGAGAGAGGTGTACTGAGCTGGGTGGCGCCGCAAGGGCATCGCGGCAAC 2307
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2308 GAGCAGATCCGACAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTCGACGGCATC 2367
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db 721 Asp 721

RESULT 4
US-10-634-165-9
; Sequence 9, Application US/10634165
; Publication No. US20050095581A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Nancy T
; APPLICANT: Gallo, Robert C
; APPLICANT: Wong-Staal, Flossie
; TITLE OF INVENTION: DETECTION OF HIV-1 DNA
; FILE REFERENCE: 223695
; CURRENT APPLICATION NUMBER: US/10/634,165
; CURRENT FILING DATE: 2003-08-05
; PRIOR APPLICATION NUMBER: 08/463,028
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/693,866
; PRIOR FILING DATE: 1985-01-23
; PRIOR APPLICATION NUMBER: 06/659,339
; PRIOR FILING DATE: 1984-10-10
; PRIOR APPLICATION NUMBER: 06/643,306
; PRIOR FILING DATE: 1984-08-22
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 9
; LENGTH: 1015
; TYPE: PRT
; ORGANISM: T cell leukemia-lymphoma virus (HTLV)
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1015)
; OTHER INFORMATION: /note= "pol protein of HTLV-III"
US-10-634-165-9

Alignment Scores:
Pred. No.: 1.52e-163 Length: 1015
Score: 3513.00 Matches: 658
Percent Similarity: 94.27% Conservative: 33
Best Local Similarity: 89.77% Mismatches: 26
Query Match: 76.65% Indels: 16
DB: 17 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-634-165-9 (1-1015)
QY 220 TTCTTCCGCGGAGGACCTGGCTTCCCGGAGGAGCCCGCGAGTTC----- 267
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
```

```
QY 268 -----CCGAGCGAGCAAGACCGCGCCCAACAGCCCAACAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGGGCGCAACAACCCCGCAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGGCACCCCTG-----AACTTCCCCAGATCACCTGTGGCAGCGCCCTCGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGGTGGCGCGCCAGATCAAGAGGCGCCCTGTGGACACCGCGCGCGCAGCACCGTG 471
Db 81 IleIysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGACCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120
QY 532 TTCAATCAAGGTGGCGCGCAGTACGACCATCTGATCGAGATCTGCGCGCAAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGGCGCCACCCCGTGAACATATCGCGCGCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCAACCTTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTGGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY 712 CCGGCGATGACGGCCCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCATCTCGGAGGAGATGGAAGAGGGCAAGATCATCAAGATCGGCGCCCGAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
QY 832 AACCCCTACAAACCCCGCTGTCGCATCAAGAAGAGCAGCACCACCAAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGGTGGACTTCCGAGCTGAACAGCGCACCCAGCATCTTCGGAGGTGCAGCTGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCGCCACCCCGCGCTGAAGAAGAAGAGCGTGACCGTGTGGACGTGGGCGAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTCAGCTGCGCTGGACGAGGACTTCGCAAGTACACCGCTTTCACCATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACACGAGACCCCGCGCATCCGCTACCACTAGCAACGTGTGCGCCCGAGGCTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGGACGCCCAAGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCGAGATCGTGTACTACCAATGATGACGACCTGTGCTGGCGCAGCGACCTG 1251
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
QY 1252 GAGATCGGCGACGACCCGCGCAGATCGAGAGCTGGCAGACACCTGCTGCGCTGGCGC 1311
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuLeuArgTrpGly 380
```

RESULT 5

```
QY 1312 TTCACACCCCGCAAGAAGACCAAGAGGAGCCCGCTTCTGTGGATGGGCTACGAG 1371
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1372 CTGCACCCCGCAAGTGGACCGTGCAGCCCATCAGCTCCCGCAGAGGAGAGCTGACCC 1431
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1432 GTCAACGACATCCAGAAAGTGTGGGCAAGCTGAATGGGCGCACCGCAGATCTACCCCGGC 1491
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
QY 1492 ATCAAGGTGGCGCAGCTGTGCAAGCTGCTCGCGCGCCCAAGCCCTGACCGACATCGTG 1551
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1552 CCCTCACCAGGAGGCGCAGCTGGAGCTGGCGCAGAACCGCGAGATCTCTCGCGGAGCCC 1611
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1612 GTGCACGGGTGTACTACGACCCCAAGGACCTGTGTGGCGCAGATCCAGAGAGCAGGGC 1671
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1672 CACGACCACTGGACCTTACCAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAG 1731
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnProProPheLysAsnLeuLysThrGlyLys 520
QY 1732 TAGCCCAAGATGGCGACCGCCACACCAAGCAGCTGAGCAGCTGACCGGCGGTGCAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1792 AAGATCGCATGGAGAGCATCTGTGCGCAAGACCCCAAGTTCGCTCGCCCATC 1851
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1852 CAGAAGGAGACCTGGGAGACCTGTGCGACCGACTACTGCGAGCGCCACCTGGATCCCGCAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1912 TGGGAGTTGTGAAACACCCCGCTGTGTGAGCTGTGTACCGCTGGAGAGGAGGCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1972 ATCATCGCGCGCCGAGACCTTCTACGTGGACCGCGCGCCCAACCGCGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2032 AAGCGCGCTTACGTGACCGACCGCGCGCGCGCAGAGATCGTGAGCTGACCGACCCACC 2091
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2092 AACCAAGACCGCAGCTGCAGGCGCATCCAGCTGGCGCTGCAGCAGCAGCGCAGGAGTG 2151
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2152 AACATCGTACCCACACAGCAGTACGCGCTGGGCGCATCATCCAGCGCCAGCCCGCAAGAGC 2211
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2212 GAGAGCGAGCTGTTGAACCAAGATCATCGAGCAGCTGATCAAGAGGAGAGGTTGATCTG 2271
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2272 AGCTGGGTGCCCGCCCAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2332 AAGGGCATCCGCAAGTGTGTTCTCTGGACGCGCATCGAT 2370
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

US-10-093-953A-26  
; Sequence 26, Application US/10093953A  
; Publication No. US20040105871A1  
GENERAL INFORMATION:  
; APPLICANT: Robinson, Harriet L.  
; APPLICANT: Smith, James M.  
; APPLICANT: Rua, Jian  
; APPLICANT: Moss, Bernard  
; APPLICANT: Amara, Rama  
; APPLICANT: Wyatt, Linda  
; APPLICANT: Earl, Patricia  
; APPLICANT: Earl, Patricia  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING  
; TITLE OF INVENTION: AN IMMUNE RESPONSE  
; FILE REFERENCE: 12804-005002  
; CURRENT APPLICATION NUMBER: US/10/093,953A  
; CURRENT FILING DATE: 2002-03-08  
; PRIOR APPLICATION NUMBER: US 60/186,364  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/251,083  
; PRIOR FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: US 09/798,675  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: PCT/US01/06795  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: US 60/324,845  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/325,004  
; PRIOR FILING DATE: 2001-09-26  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 26  
; LENGTH: 739  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: protein encoded by construct of vaccine vector  
; OTHER INFORMATION: pGA2 and insert JS2 expressing ciade HIV-1 VL  
US-10-093-953A-26  
Alignment Scores:  
Pred. No.: 3,45e-162 Length: 739  
Score: 3485.00 Matches: 654  
Percent Similarity: 94.00% Conservative: 35  
Best Local Similarity: 89.22% Mismatches: 28  
Query Match: 76.04% Indels: 16  
DB: 16 Gaps: 3  
US-09-610-313B-30 (1-2469) x US-10-093-953A-26 (1-739)  
Qy 220 TTCTTCGGCAGGACCTGGCTTCCCGCAGGCGCAAGCGCCGAGTTCCCGCAGGAGCAG 279  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
Qy 280 AACCGC-----GCAACAGCCCAACAGC 303  
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40  
Qy 304 CGCGAGCTGAGGTG-----CGCGCGACACCCCGCGAGCGAGCGCGCGCGCGCGCGC 357  
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60  
Qy 358 CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGC 411  
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80  
Qy 412 ATCAAGGTGGCGCGCAGATCAAGGAGGCCCTGTGGACACCGCGCGCGCGCAGCACCGTG 471  
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100  
Qy 472 CTGAGGAGATGAGCTGCCCGGCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGC 531  
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGly 120

```
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuileAlaGluileGlnLysGlnGly 500
Qy 1672 CAGACACAGTGGACCTACAGATCTACAGAGCCCTTCAAGAACCTGAAGACGGCAAG 1731
Db 501 GlnGlyGlnTrpThrTyrGlnileTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
Qy 1732 TAGCCCAAGATGGCAGCCGCCACCAACAGCTGAAGCAGCTGACCGAGCGCGTGCGAG 1791
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
Qy 1792 AAGATCCCATGGAGACATCGTATCTGGGGGAAGACCCCAAGTTCCGCTGCCCATC 1851
Db 541 LysileThrThrGluSerileValileTrpGlyLysThrProLysPheLysLeuProile 560
Qy 1852 CAGAAGAGACCTGGGAGACCTGTGGACCGACATCTAGGAGCGCCACCTGGATCCCGAG 1911
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
Qy 1912 TGGGAGTTCGTGAACCCCGCTGCTGAAGCTGTGTACAGCTGGAGAGGAGGCC 1971
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpLysGlnLysGluPro 600
Qy 1972 ATCATCGCGCCGAGACCTTCTAGTGGACGGCGCCCAACCGCGAGACCAAGATCGGC 2031
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
Qy 2032 AAGCCGGCTAGTGACCGACCGCGCGCGGAGAGATCGTGAAGCTGGACCGACGACC 2091
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
Qy 2092 AACCAAGACCGAGCTGCGAGGCCATCCAGCTGCGCTGCGAGACAGCGCGAGGTG 2151
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
Qy 2152 AACATCTGTGACGACACCGCTAGTCCGCTGGCATCATCAGGCGCCAGCCCGCAAGAGC 2211
Db 661 AsnileValThrAspSerGlnTyrAlaLeuGlyileileGlnAlaGlnProAspLysSer 680
Qy 2212 GAGACGAGCTGTGAACACGATCATCGACGATGATCAAGAGGAGAGGTGTACCTG 2271
Db 681 GluSerGluLeuValAsnGlnileileGluGlnLysLysGluLysValTyrLeu 700
Qy 2272 AGCTGGGTGCGCGCCACAGGCGATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAGC 2331
Db 701 AlaTrpValProAlaHisLysGlyileGlyGlnGlnGlnValAsnGluGlnValAspLysLeuValSer 720
Qy 2332 AAGGCGATCGCGAAGGTGTCTTCTCGGCGGATCGAT 2370
Db 721 AlaGlyileArgLysileLeuPheLeuAspGlyileAsp 733
```

## RESULT 6

```
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE OF INVENTION: AN IMMUNE RESPONSE
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093,953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
```

```

; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by construct of vaccine vector
; OTHER INFORMATION: pGAL and vaccine insert expressing clade B HIV-1
; OTHER INFORMATION: Gag-Pol
; US-10-093-953A-31

Alignment Scores:
Pred. No.: 3 45e-162 Length: 739
Score: 3495.00 Matches: 654
Percent Similarity: 94.00% Conservative: 35
Best Local Similarity: 89.22% Mismatches: 28
Query Match: 76.04% Indels: 16
DB: Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-093-953A-31 (1-739)
Qy 220 TTCTTCCGAGAGACTGCGCTTCCCTCCAGGGGAAGCCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerGluGln 20
Qy 280 AACCGC-----CCCAACAGCCCCACGAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40
Qy 304 CGCGAGCTGCAGGTG-----CGCGCGGACCAACCCCGCAGCGAGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
Qy 358 CAGGCGACCCCTG-----AACTTCCCTCCAGATCACCCTGTGGCAGCGCCCTCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnileThrLeuTrpGlnArgProLeuValThr 80
Qy 412 ATCAAGGTGGCGCGCGCAGATCAAGGAGGCGCTCTCGACACCGCGCGCGCGCGCACCGCTG 471
Db 81 IleLysileGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
Qy 472 CTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetileGlyGlyileGlyGly 120
Qy 532 TTCATCAAGGTGGCGCGCAGTACGACCGAGATCCTGATCGAGATCTGCGGCAAGAGCGCATC 591
Db 121 PheileLysValArgGlnTyrAspGlnileLeuileGluileCysGlyHisLysAlaile 140
Qy 592 GGCACCGTGTGATCGCGCGCCCGCGCGTGAACATCATCGCGCGCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnileleGlyArgAsnLeuLeuThrGln 160
Qy 652 CTGGGTGTCACCTGAACTTCCCATCAGCCCATCGAGACCGCTGCGCGCTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProileGluThrValProValLysLeuLys 180
Qy 712 CCGCGCATGAGCGCGCCCAAGGTGAAGAGTGGCCCTGACCGGAGGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysileLysAla 200
Qy 772 CTGACCGCCCATCTGCGGAGGAGATGGAGAGGAGGAGATCACCAGATCGCGCGCGAG 831
Db 201 LeuValGluileCysThrGluMetGluLysGluLysileSerLysileGlyProGlu 220
```

QY 832 AACCCCTACACACCCCGCTGTCGCCATCAAGAGAGGACAGACCAAGTGGCGAAG 891  
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240  
QY 892 CTGGTGGACTTCCGGAGCTGAACAAGCGCACCCAGAGACTTCTGGAGGTGCAGCTGGCC 951  
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260  
QY 952 ATCCCCACCCCGCGGCTCAAGAAGAAGAGCGTGAACCGTGTGGAGCTGGCGAAC 1011  
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280  
QY 1012 GCCTACTTCAGCGTGGCCCTGGAGAGACTTCGCAAGTACAGCGCTTCACCATCCCC 1071  
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTrpAlaPheThrIlePro 300  
QY 1072 AGCATCAACACAGAGACCCCGGATCCGCTACAGTACAACTGCTGCCCGCAGGCTGG 1131  
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320  
QY 1132 AAGGGCAGCCCGACATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTCCGCGCC 1191  
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340  
QY 1192 CGCAACCCCGAGATCGTGATCTACCGTACATGACGACCTGTAGTGGCGAGCGACTG 1251  
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360  
QY 1252 GAGATCGGCCAGCACCGGCCAAGATCGAGAGCTGGCAAGCACTGCTCGCTGGGGC 1311  
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380  
QY 1312 TTCAACCCCGCACAAGAAGCACCAAGAGAGCCCGCTTCTGTGTGATGGGTACGAG 1371  
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400  
QY 1372 CTGCACCCCGACAGTGGACCGCTGAGCCCATCGAGTGGCCGAGAGAGAGCTGGACC 1431  
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420  
QY 1432 GTGAACACATCCAGAGCTGCTGGCAAGCTGAACCTGGGCGACGAGATCTACCCCGGC 1491  
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440  
QY 1492 ATCAAGTGGCCAGCTGTGAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCGTG 1551  
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460  
QY 1552 CCCCTGACCGAGGAGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCGCGGAGGCC 1611  
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480  
QY 1612 GTGCACCGCTGTACTACGACCCAGCAAGACCTGTTGGCGGAGATCCAGACGAGGC 1671  
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500  
QY 1672 CACGACCACTGGACCTACCACTATCCAGAGAGCCCTTCAAGAACCTGAAGACCGCAAG 1731  
Db 501 GlnGlyGlnThrThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520  
QY 1732 TACGCCAAGATGCGCAGCGCCCAACCAACGACGTGAAGCAGCTGACGAGCGCGTGGAG 1791  
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540  
QY 1792 AAGATCCCATGGAGACATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTCCCATC 1851  
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560  
QY 1852 CAGAAGGAGACCTGGGAGACCTGTGGACCACTACTGCGAGGCGCACCTGATCCCCGAG 1911  
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580  
QY 1912 TGGAGTTGCGTAACACCCCGCTGTTGAGCTGTGGTACCAGCTGTGGAGAGGAGGCC 1971

Db 581 TrpGluPheValAsnThrProLeuValLysLeuTyrTrpGlnLeuGluLysGluPro 600  
QY 1972 ATCATCGCGCGCAGAGACTTCTACGTGGACCGCGCCCGCAACCGCGAGACCAAGATCGGC 2031  
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620  
QY 2032 AAGCGCGCTACGTGACGACCGCGCGCGCGAGATCTGTAGCCTGTACCGAGACCAACC 2091  
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640  
QY 2092 AACCAAGAGCCGAGCTGCAGGCCATCCAGCTGCGCTTCGAGGACAGCGCGAGGAGTG 2151  
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660  
QY 2152 AACATCTGTGACCGACAGCCAGTACGCCCTGGGCGATCATCCAGGCCCGACCAAGAGC 2211  
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680  
QY 2212 GAGAGCAGCTGTGTGAACCGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTG 2271  
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700  
QY 2272 AGCTGGGTGCGCGCCCAAGGCGCATCGCGGCAACGAGCAGATCGACAACCTGCTGAGC 2331  
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGlnValAspLysLeuValSer 720  
QY 2332 AAGGCGATCCGCAAGTGTCTTCTCGACGCGCATCGAT 2370  
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7  
US-10-325-468-23  
; Sequence 23, Application US/10325468  
; Publication No. US20040101823A1  
; GENERAL INFORMATION:  
; APPLICANT: Soong, Nay Wei  
; APPLICANT: Pekrun, Katja  
; APPLICANT: Shibata, Riri  
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
; FILE REFERENCE: 0166.210US  
; CURRENT APPLICATION NUMBER: US/10/325,468  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US 60/343,524  
; PRIOR FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 56  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 1003  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol  
US-10-325-468-23

Alignment Scores:  
Pred. No.: 2,38e-161 Length: 1003  
Score: 3468.00 Matches: 642  
Percent Similarity: 95.28% Conservative: 45  
Best Local Similarity: 89.04% Mismatches: 30  
Query Match: 75.67% Indels: 4  
DB: 16 Gaps: 2

US-09-610-313B-30 (1-2469) x US-10-325-468-23 (1-1003)

QY 220 TTCTTCGCGAGAGACCTTGGCTTCCCCAGGCGCGCGAGTTCCTCCAGCGAGCAG 279  
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20  
QY 280 AACCGCGCCCAACGCCCGCCAGCCGCGAGCTGCGAGTG-----CGCGGCGACAACCCC 333  
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40

QY 334 CGCAGCGAGCGCGCGCGAGCGCGCAGCGCACCTG-----AACTTCCCGCCAGATCAC 387  
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnProGlnIleThr 60  
QY 388 CTGTGGCAGGCGCCCTGTGTGAGCATCAAGGTGGCGGCAGATCAAGAGAGCGCTGCTG 447  
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
QY 448 GACACCGCGCGGACGACCGTGTGTGAGAGATGAGCTGCCCGGCAAGTGAAGCCC 507  
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100  
QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCAGTACGACACGATCCTGATC 567  
Db 101 LysMetIleGlyIleGlyIleGlyPheIleLysValArgGlnIleAspGlnIleProIle 120  
QY 568 GAGATCTGCGCGCAAGAGCGCATCGGCACCGCTGATCGCGCCCGCCCGCGTGAACATC 627  
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140  
QY 628 ATCGCGCGCAACATGTCAGCCAGCTGGGTGTGACCTGAACTTCCCGCATCGCCCATC 687  
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
QY 688 GAGACCGTGCCTGTGAAGCTGAAGCGCGCATCGACGCGCCCGCAAGGTGAAGCAGTGGCCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGGAGAGAGATCAAGGCGCTGACCGCCATCTCGCAGGAGATGGAAGAGAGGCG 807  
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AAGATCACCAAGATCGCGCGCGAGAACCTCTACAACACCCCGTGTTCGCCATCAAGAAG 867  
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
QY 868 AAGSACAGCACCAAGTGGCGCAGCTGGTGACTTCGCGCAGCTGAACAGCGCACCCAG 927  
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240  
QY 928 GACTTCTGGAGGTGTCAGCTGGGCATCCCCACCCCGCGCGCTGAAGAAAGAAAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
QY 988 GTGACCGTGTGACGTGGCGCGCGCTACTTACGCGTGGCGCGCTGGCAGGACTTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCGCTTACCATCCCGCAGCATCAACAGCAGACCCCGCGCATCCGCTACCAG 1107  
Db 281 LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 300  
QY 1108 TACAACGTGTGCGCCCGAGGCTGGAAGGCGAGCGCCACGATCTTCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTGTGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGTATCTACAGTACATGGAC 1227  
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleTyrGlnTyrMetAsp 340  
QY 1228 GACTGTACCTGGCGCAGCTGGAGATCGGCACCGCAGCCCGCGCGAGATCGAGGAGCTG 1287  
Db 341 AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1288 CGCAAGCACCTGTGCTGCGGCTTCCACCAACCCCGCACAAGAGACCAAGAGAGAGCCC 1347  
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1348 CCTTCTGTGGATGGGCTACGAGCTGCACCCGACCAAGTGGAGCGGTGAGCGCCATCGAG 1407  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpValGlnProIleVal 400

QY 1408 CTGCGCGAGAGGAGAGCTGGACCGGTGAACGACATCCAGAGAGCTGGTGGCGAAGCTGAAC 1467  
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1468 TGGCGCAGCGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGAAGCTGTGCGCGGC 1527  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440  
QY 1528 GCCAAGCGCCTGACCCGACATCGTGGCCCTGACCGCAGGAGCGGAGCTGGAGCTGGCGCGAG 1587  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuAlaGlu 460  
QY 1588 AACCGCGCAGATCTCGCGCAGCGCGCTGTACTACGACCCCGCAGAGAGAGCTG 1647  
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480  
QY 1648 GTGGCGCAGATCCAGAAGCAGGCGCCACGACCTGAGTGGACCTTACAGATCTACAGGAGCCC 1707  
Db 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500  
QY 1708 TTCAAGAACTGAAGACCGCGCAAGTACGCCAAGATCGGCACCGCCACACCAACGACGCTG 1767  
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgSerAlaHisThrAsnAspVal 520  
QY 1768 AAGCAGCTGACCGAGCGCGCTGCAGAAAGATCGCCATGGAGAGCATCGTGTCTGGGCAAG 1827  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540  
QY 1828 ACCCCAAAGTTCCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887  
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560  
QY 1888 TGGCGCGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTG 1947  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580  
QY 1948 TGGTACAGCTGGAAGAGGAGGCCATCATCGCGCGCAGACCTTCTACGTGAGCGCGCC 2007  
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
QY 2008 GCCAACCGCGCAGACCAAGATCGGCAGGCGCGCTAGTGAACCGACCGCGGCGCGCAGAG 2067  
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620  
QY 2068 ATCGTGAAGCTGACCGCAGACCCACCAAGAGACCGAGCTGCAGCGCATCCAGCTGGCC 2127  
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
QY 2128 CTGCAGCAGCGCGCAGCGAGGTGAACATCGTGACCGCAGCAGCGAGTACGCCCTGGGCATC 2187  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
QY 2188 ATCAGGCGCGCAGCCCGCAAGCAGCAGGAGCGAGCTGGTGAACAGATCATCGAGCAGCTG 2247  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
QY 2248 ATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCGCGCCACCAAGGCGCATCGCGCGCAAC 2307  
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700  
QY 2308 GAGCAGATCGACAAGCTGGTGAAGCAAGGCGCATCCGCAAGGTGTGTCTCTGGACGGCATC 2367  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720  
QY 2368 GAT 2370  
Db 721 Glu 721

## RESULT 8

US-10-325-468-35  
; Sequence 35, Application US/10325468  
; Publication No. US20040101823A1  
; GENERAL INFORMATION:

```

; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol
; US-10-325-468-35

Alignment Scores:
Pred. No.: 2,98e-161 Length: 1003
Score: 3466.00 Matches: 643
Percent Similarity: 95.28% Conservative: 44
Best Local Similarity: 89.18% Mismatches: 30
Query Match: 75.63% Indels: 4
DB: 16 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-325-468-35 (1-1003)

Qy 220 TTCTTCGGCAGGACCTGGCTTCCCGCAGGGCAAGCGCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerSerGluGln 20
Qy 280 AACCGCGCCCAACGCCCCACCAGCGCGAGCTGCAGGTG---CGCGCGAC---AACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgGlyAspAsnSer 40
Qy 334 CGCAGCGAGGCGCGCGCGCGAGCGCGAGCGCACCTCGT-----AACTTCCCGCCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
Qy 388 CTGTGGCAGGCGCCCTGTGTGAGCATCAAGTGGCGCGCGCAGATCAAGAGGAGCGCTCTGT 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCGACACCGTCTCGAGGAGATGAGCTGCCCGCGCAAGTGGAGCGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
Qy 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGCGCGAGTACGACGACGATCCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle 120
Qy 568 GAGATCTGGCGCAAGAGGCGCATCGCACCGTGTGATCGGCGCCCGCCCGCGCGCAATC 627
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
Qy 628 ATCGCGCGCAACATGTCGACCGCGCTGCGCGTGCACCTGAACTTCCCGCATCGCGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
Qy 688 GAGACCGTGCCTGAACTGAAAGCGCGCGATGACCGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
Qy 748 CTGACCGGAGGAGATCAAGGCGCTGACCGCCATCTCGCAGGAGATGGAAGAGGAGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
Qy 808 AAGATCACCAAGATCGCGCGCGCGAGAACCTCTACAGACCGCGCGCTGTCGCGCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220

```

```

Qy 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGGAGTTCCTCCGCGAGCTGAAACAACGCGACCCGAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240
Qy 928 GACTTCTGGGAGGTGACGTGGGCATCCCGCCACCCCGCGCGCTGAGAGAAAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLys 260
Qy 988 GTGACCGTGTGGACGTGGCGCGACCTACTTACGCTGCGCCCTGGAGCAGGACTCTCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
Qy 1048 AAGTACACCGCTTCCACATCCCGCAGCATCAACAACGAGACCCCGCGCATCCGCTACCGAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
Qy 1108 TACACGCTGTGCTCCCGAGGCTGGAAGCGGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
Qy 1168 AAGATCCTGGAGCGCTTCCCGCGCGCGCAACCCCGAGATCGTATCTACAGTACATGAGC 1227
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
Qy 1228 GACTGTATCTGGGACGACCTGGAGATCGCGCAGCACCGCGCAAGAGATCGAGAGCTG 1287
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
Qy 1288 GCGAAGCAGCTGTGCTGGCGCTTCCACCGCGCGCGAGATCGTATCTACAGTACATGAGC 1347
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
Qy 1348 CCCTCTCTGTGGATGGGTACGAGCTGACCGCGCAAGTGGACCGTGGAGCCCATCGAG 1407
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
Qy 1408 CTCCCGGAGAGGAGCTGACCGTGAACGACATCCAGATCCAGAGCTGTGGGCAAGCTGAAC 1467
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy 1468 TGGCGCGAGCATCTACCCCGCGCTCAAGTGGCGCGCGAGCTGTGCAAGCTGTGCGCGGC 1527
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLysLysLysLeuLeuArgGly 440
Qy 1528 GCCAAGCGCTGACCGACATCGTGCCTCCACGAGGAGCGCGAGCTGGAGCTGGCGGAG 1587
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuLysLysLys 460
Qy 1588 AACCGCGAGATCTGCGCGAGCGCGTGCACCGCGGTGTACTACGACCCCGCAGAGAGCTG 1647
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480
Qy 1648 GTGGCGGAGATCCAGAGCAGGCGCCACGACGATGGAGCTACCGATCTACCGAGCGCC 1707
Db 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500
Qy 1708 TTCAAGAACCTGAAAGCGCGCAAGTACGCCAAGATGGCGCGCGCGCGCGCGCGCGCGCG 1767
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
Qy 1768 AAGCAGCTGACCGAGGCGGTGCAGAAAGATCCCATGGAGAGCATCGTGTGATCTGGGGCAAG 1827
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
Qy 1828 ACCCGCAGGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1887
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
Qy 1888 TGCAGCGCCACCTGGATCCCGCGAGTGGAGTTCGTGAACACCCCGCGCTGGTGAAGCTG 1947
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

```



```
QY 1948 TGGTACCAGCTGGAGAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCC 2007
Db 581 TTPtyrGlnLeuGluLysGluProlIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCGAGACCAAGATCGCAAGCGCGGTACGTGACCCAGCCGGGCGCGCAGAAG 2067
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
QY 2068 ATCGTGAAGCTGACCGAGACCAACCAAGACCGAGTGGAGCCATCCAGCTGGCC 2127
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2128 CTCAGGACAGCGCGAGCGAGGTGAACATCGTGACCGACAGCGAGTACGCGCTGGGCATC 2187
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCCCGCCGACCAAGAGCGAGCGAGCTGGTGAACAGATCATCGAGCGCTG 2247
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCGCCGCCACCAAGGGCATCGCGGCAC 2307
Db 681 IleLysLysGluLysValTyrLeuThrTrpIleProAlaHisLysGlyIleGlyGlyAsn 700
QY 2308 GAGCAGATCGACAGCTGGTGAGCAAGGGCATCGCAAGGTGCTGTCTCTGGACGGCATC 2367
Db 701 GlnGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db 721 Glu 721
```

## RESULT 9

```
US-10-325-468-9
; Sequence 9, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Shibata, Riri
; APPLICANT: Pekrun, Katja
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; FILE REFERENCES: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;
; OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;
; OTHER INFORMATION: clone P10.26 protein Pol
US-10-325-468-9
```

## Alignment Scores:

Pred. No.:	3.33e-161	Length:	1003
Score:	3465.00	Matches:	642
Percent Similarity:	95.15%	Conservative:	44
Best Local Similarity:	89.04%	Mismatches:	31
Query Match:	75.61%	Indels:	4
DB:	16	Gaps:	2

US-09-610-313B-30 (1-2469) x US-10-325-468-9 (1-1003)

```
QY 220 TTCTTCGGCAGGACCTGGCGCTTCCCCCAGGGCAGCCCGCGAGTTCCCCCAGGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
```

```
QY 280 AACCGCGCAACAGAGCCCGCCACCGAGCGCGAGCTGCAGGTG-----CGCGCGCAACACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTipArgArgAspAsnAsnSer 40
QY 334 CGAGAGAGCGCGCGCGCGCGAGCGCCAGGGCACCCCTG-----AACTTCCCCAGATCACCC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCGCCCTCGTGTGAGCATCAAGTGGGGCGGCAGATCAAGAGAGGCCCTGTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGGGCGCCGACACACCGCTGTGTGAGAGAGATGAGCTGCCCGCAAGTGAAGACCCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTyrLysPro 100
QY 508 AGATGATCGGGCGCATCGCGCGCTTCATCAAGGTGGCGCAGTACACCGAGTCTCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120
QY 568 GAGATCTGGCGCAAGAGGCCCATCGGCACCGCTGTGATCGGCCCGCCACCGCTGAACATC 627
Db 121 AspileCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGTGACCCAGCTGGGCTGCACCCCTGAACCTTCCCATCAGCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGAAGCTGAAGCCCGCGCATGGACGGCGCCCGCCCAAGGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAAGAGAGGGGC 807
Db 181 LeuThrGluGlnLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCCCGCGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AGGACAGACCAAGTGGCGCAAGCTGTGTGGTTCCTCGCGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240
QY 928 GACTCTTGGGAGGTGCAGCTGGGCATCCCCACCGCGCGCTGAAGAGAGAAAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGGACGTGGCGCGACGCTACTTTCAGCGTGCCTGAGCAGAGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AGTACACCGCTTCACCTCCCGCATCAACAGAGACCCCGCGCATCCCGCTACCGAG 1107
Db 281 LysThrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAAGTGTGTCGCCAGCGCTCGAAGGCGCAGCCAGCATCTTCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTGTGAGCCCTTCCGCGCGCCGCAACCCCGAGATCGTGTATCCAGTTCATGGAC 1227
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
QY 1228 GACCTGTAGTGGCGAGCGACCTGGAGATCGGCACGACCGCCCGCAGATCGAGGAGCTG 1287
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1288 CGCAAGCACCTGTGCTGGGCTTCACCCACCGCCGACAAAGAGACCAAGAGAGAGAGGCC 1347
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCCTTCTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGACCGTGCACCCATCGAG 1407
```

```
|||||
381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY CTGCCGAGAGAGAGCTGACCGTGAACACATCCAGAACCTGGTGGGCAAGCTGAAC 1467
Db |||||
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY TGGCCAGCCAGATCTTACCCCGGATCAGTGTGGCCAGCTGTCTGCTCGCGGC 1527
Db |||||
421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
QY 1528 GCCAAGCCCTGACCGCATCGTCCCTGACCCAGGAGGCGAGCTGGAGCTGCCGAG 1567
Db |||||
441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460
QY 1588 AACCGCAGATCTCTGCGCAGCCCGTGCACGCGCTGTACTACGACCCAGCAAGACCTG 1647
Db |||||
461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1648 GTGCCCGAGATCCAGAGCGAGGCGCACGACAGTGGACCTTACGATCTACGAGGCC 1707
Db |||||
481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500
QY 1708 TTCAAGAACCTGAGACCGCGCAAGTACGCCAAGATGCGCACCGCCACCAACGACGTG 1767
Db |||||
501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGCGCTGCAGAAATCCCATGGAGAGCATCTGTGATCTGGGGCAAG 1827
Db |||||
521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
QY 1828 ACCCCCAAGTTCGCCCTGCCCATCCAGAGAGACCTGGGAGACCTGGTGACCGACTAC 1887
Db |||||
541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1888 TGGCAGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTG 1947
Db |||||
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1948 TGGTACCAGCTGGAGAGAGCCCATCATCGGCCCGAGACCTTCTACGTGGAGCGCGCC 2007
Db |||||
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCAGACCAAGATCGCAAGCGCGCTTACGTGACCGACCGCGCGCGCAGAG 2067
Db |||||
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
QY 2068 ATCGTGACCTGACCGAGACCAACCAAGAGACCGAGCTGCGAGGCTATCCAGCTGGCC 2127
Db |||||
621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2128 CTGCAGCAGACCGCGCAGGAGTGAACATCGTGACCGACAGCAGTACGCCCTGGGCATC 2187
Db |||||
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCAGCCCGCAGCCGCAAGAGCAGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTG 2247
Db |||||
661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2248 ATCAGAGAGAGAGGTGTACTGAGCTGGTGGTGGCCCGCCCAAGGGCATCGCGCGCAAC 2307
Db |||||
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
QY 2308 GAGCAGATCCACAAGCTGGTGAGCAAGGGCATCGCAAGGTGTGTCTCGGAGCGGCATC 2367
Db |||||
701 GluGlnValAspLysLeuValSerAlaGlyIleArgValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
Db |||||
721 Glu 721
Db |||||
RESULT 10
US-10-325-468-20
```

```
; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; PRIOR FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20
```

```
Alignment Scores:
Pred. No.: 3,33e-161 Length: 1003
Score: 3465.00 Matches: 642
Percent Similarity: 95.15% Conservative: 44
Best Local Similarity: 89.04% Mismatches: 31
Query Match: 75.61% Indels: 4
DB: 16 Gaps: 2
```

US-09-610-313B-30 (1-2469) x US-10-325-468-20 (1-1003)

```
QY 220 TTCTTCGCGAGGACCTGGCTTCCCCAGGCGAAGCCCGCGAGTCTCCCGAGGAGCAG 279
Db 1 PhePheargGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
QY 280 AACCGCCCAACAGCCGCCACCGCGAGCTGCAGGTG-----CGCGGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProIleArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCGAGCGCACCTG-----AACTTCCCGCATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCGCTGGTGAGCATCAAGGTGGCGCGCAGATCAAGAGGAGCGCTGTCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCACACCGTCTCTCGAGGAGATGAGCTGCGCGCAAGTGAAGGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGCTTTCATCAAGGTGGCGCGAGTACGACCGAGATCTGTATC 567
Db 101 LysMetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120
QY 568 GAGATCTGCGCGCAAGAGCGCATCGCACCGCTGTGATCGCGCCCGCGCGCGTCAACATC 627
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGACCCAGCTGGGTGCACTGAACTTCCCATCAGCCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGGCGCTGAGCTGAGCCCGGATGAGCCCGGATGAGCCCGCGCGCGTGGAGCTG 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGGAGGAGATCAAGGCGCTGACCGCCCATCTGCGAGGAGATGGAGAGAGGCG 807
Db 181 LeuThrGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
```



Db 21 ThrGlyAlaAenSerSerAlaSerArgLysLeu-----GlyAspGly----- 34  
Qy 340 GAGCGCGCGCGGAGCGCGAGCGGACCC-----CTGAATCTTCCCCCAGATCACCCCTG 390  
Db 35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheGlnIleThrLeu 53  
Qy 391 TGGCAGGCGCCCTTGGTGGAGCATCAAGGTGGGGCGGCGAGATCAAGAGGCGCCCTGCTGAC 450  
Db 54 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 73  
Qy 451 ACCGCGCGCGAGCACCGCTGGAGGAGATGAGCTGCCCGGCAAGTGAAGCCCAAG 510  
Db 74 ThrGlyAlaAspPheThrValLeuGluAspIleAsnLeuProGlyLysIleThrLysProLys 93  
Qy 511 ATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGCGGAGTACGACCGACCATCTGATCGAG 570  
Db 94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 113  
Qy 571 ATCTGGCGCAAGAGCGCATCGGCACCGTGTGATCGGCGCCCGACCCCGTGAACATCATC 630  
Db 114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133  
Qy 631 GGCGCAACATGTCGACCCAGCTGGCTGACCTGACCTTCCCATCAGCCCATCGAG 690  
Db 134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153  
Qy 691 ACCGTGCGCGTGAAGTGAAGCCCGCATGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 750  
Db 154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173  
Qy 751 ACCGAGGAGAGATCAAGCGCCCTGACCGCCATCTCGGAGGAGATGGAAGAGGCGCAAG 810  
Db 174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluLys 193  
Qy 811 ATCACCAGATCGCGCGCGGAGAACCCCTAGAACACCCCGCTGTTCCGCGCATCAAGAGAG 870  
Db 194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 213  
Qy 871 GACAGCACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTTGAACAGCGCACCGAGGAC 930  
Db 214 AspSerThrLysThrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233  
Qy 931 TTCTGGAGGTGACAGCTGGGCATCCCCACCCCGCGCTTGAAGAGAGAGAGCGGTG 990  
Db 234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 253  
Qy 991 ACCGTGTGAGCTGGGGAGCGCTACTTACGCTGCGGCGGCGGAGGAGGACTTCCGCAAG 1050  
Db 254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273  
Qy 1051 TACACCGCTTTCACCATCCCGAGCATCAACAGAGACCCCGCGCATCCGCTACCGTAC 1110  
Db 274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 293  
Qy 1111 AACGTGTGCGCGCGGCTGAAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1170  
Db 294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313  
Qy 1171 ATCTGAGGCGCTTCCGCGCGCGGCAACCCCGAGATCGTATCTACCGATCATCGAGCAC 1230  
Db 314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333  
Qy 1231 CTGTACCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1290  
Db 334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353  
Qy 1291 AAGCACCTGCTGCGCGGCTTCAACACCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1350  
Db 354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373  
Qy 1351 TTCTGTGGATGGCTACGAGCTGCACCCGACCAAGTGGACCGTGGACCGCATCGAGCTG 1410

Db 374 PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393  
Qy 1411 CCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGAGCTGGTGGGCAAGCTGAACCTGG 1470  
Db 394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413  
Qy 1471 GCCAGCCAGATCTACCCCGCGCATCAAGGTGGCGGCGGAGCTGCTCAAGCTGCTGGCGGCGCC 1530  
Db 414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433  
Qy 1531 AAGCCCTGACCCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1590  
Db 434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuGluGluAsn 453  
Qy 1591 GCGAGATCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1650  
Db 454 ArgGluIleLeuAspGluProValHisGlyValTyrAspProSerLysAspLeuVal 473  
Qy 1651 GCCAGATCCAGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1710  
Db 474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493  
Qy 1711 AAGAACCTGAAGACCGCGAGTACGCCAAGATGCGCACCGCGGCGGCGGCGGCGGCGGCGGCGG 1770  
Db 494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513  
Qy 1771 CAGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1830  
Db 514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533  
Qy 1831 CCCAAGTTCGCTGCGCGCATCCAGAGAGAGACCTGGGAGAGCTGGTGGACCGCATCTACTGG 1890  
Db 534 ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553  
Qy 1891 CAGSCACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGCGCGCGCGCGCGCGCGCGCGCGG 1950  
Db 554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573  
Qy 1951 TACCAGCTGGAGAGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2010  
Db 574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593  
Qy 2011 AACCGCGAGACCAAGATCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2070  
Db 594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613  
Qy 2071 GTGAGCTGACCGGAGACCCACCAAGAGACCGGAGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGG 2130  
Db 614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
Qy 2131 CAGGACGCGGCGGCGGAGGTGAACATCGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2190  
Db 634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653  
Qy 2191 CAGCGCGGCGGCGGAG 2250  
Db 654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673  
Qy 2251 AAG 2310  
Db 674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693  
Qy 2311 CAGATCGACAGCTGGTGGAGCAAGGCGATCGCAAGGTGCTGCTGCTGCGAGCGGATCGAT 2370  
Db 694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713

## RESULT 12

US-10-296-734-1470  
; Sequence 1470, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A

; TITLE OF INVENTION: Synthetic molecules and uses therefor  
 ; FILE REFERENCE: Savine  
 ; CURRENT APPLICATION NUMBER: US/10/296,734  
 ; CURRENT FILING DATE: 2003-08-04  
 ; PRIOR FILING DATE: 2000-05-26  
 ; NUMBER OF SEQ ID NOS: 1507  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 1470  
 ; LENGTH: 995  
 ; TYPE: PRT  
 ; ORGANISM: Human immunodeficiency virus  
 US-10-296-734-1470

Alignment Scores:  
 Pred. No.: 3,95e-160 Length: 995  
 Score: 3443.00 Matches: 642  
 Percent Similarity: 94.72% Conservative: 40  
 Best Local Similarity: 89.17% Mismatches: 28  
 Query Match: 75.13% Indels: 10  
 DB: 15 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-296-734-1470 (1-995)

QY	220	TTCTTCGCGAGGACCTGGCTTCCCGGCAAGCCCGCGAGTTCCCGAGCGAGCAG	279
DB	1	PhePheArgGluAsnLeuAlaPheGlnGlnGlyLysAlaArgGluPheSerSerGluGln	20
QY	280	AACCGCGCCAAACGCCACCGCGGAGCTGAGGTGCGCGCGAGCAACCCCGCGAGC	339
DB	21	ThrGlyAlaAsnSerSerAlaSerArgLysLeu-----GlyAspGly-----	34
QY	340	GAGCGCGCGCGCGAGCCGCGAGGCGCACC-----CTGAACCTCCCGAGATCACCCCTG	390
DB	35	---GlyGlyAlaGluArgGlnGlyThrSerSerPheSerPheProGlnIleThrLeu	53
QY	391	TGGCAGCGCCCTCGTGTGAGCATCAAGGTGGGGCGGCAGATCAAGAGAGCGCTGCTGGAC	450
DB	54	TrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLysGluAlaLeuLeuAsp	73
QY	451	ACCGCGCCGACGACACCGTGTGGAGGAGATGAGCTGCCCGCGCAAGTGGAAGCCCAAG	510
DB	74	ThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys	93
QY	511	ATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGGCAGTACGACGAGATCCTGATCGAG	570
DB	94	MetIleGlyGlyIleGlyLysPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu	113
QY	571	ATCTGCGCGCAAGAGGCGCATCGGCACCGTGTGATCGCGCCCGCCCGCGTGAACATCATC	630
DB	114	IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle	133
QY	631	GGCGCGCAACATGTCACCCAGCTGGGTGCACCTGAACTTCCCATCAGCCCGCATCGAG	690
DB	134	GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp	153
QY	691	ACCGTCCCGCTGAGCTGAAGCCCGCATCGACCGCCCGCAAGTGTGAGAGTGGCCCTG	750
DB	154	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	173
QY	751	ACCGAGGAGAAGATCAAGGCGCTGACCGCATCTCGCGAGAGATGGAAGAGGAGGCAAG	810
DB	174	ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluLysLys	193
QY	811	ATCACCAAGATCGCGCCGAGAACCCCTACAACACCCCGCTGTTCGTCATCAAGAAGAAG	870
DB	194	IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys	213
QY	871	GACAGCACCAAGTGGCGCAGCTGTGAGTTCGCGAGCTGACAGCGCCACCCAGGAC	930
DB	214	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	233
QY	931	TTCTGGGAGGTGCAGCTGGGCATCCCCACCGCCCGCTCGAAGAAGAAGAGCGTG	990

DB	234	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	253
QY	991	ACCGTCTGGAGCTGGCGGACGCTACTTTCAGCGTCCCTCGACGAGACTTCCGCAAG	1050
DB	254	ThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluSerPheArgLys	273
QY	1051	TACACCGCTTACCATCCCGAGCATCAACAAAGAGAGACCCCGCGCATCCGCTACCAAGTAC	1110
DB	274	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr	293
QY	1111	AACGTCTGCTCCCGAGGCTGGAAGGAGCGCCCGAGCATCTTCCAGAGCAGCATGACCAAG	1170
DB	294	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	313
QY	1171	ATCCTGGAGCCCTTCGCGCGCCCAACCCCGAGATCGTGTATCTACAGATACATGGAGCAG	1230
DB	314	IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp	333
QY	1231	CTGTACGTGGGACGAGCATCGGAGATCGGACACCGCGCCCAAGATCGAGGAGCTGCGC	1290
DB	334	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	353
QY	1291	AAGCACCCTGCTGGCTGGGCTTTCACACCCCGCAACAAGAGCACCAGAGAGGCCCC	1350
DB	354	AlaHisLeuLeuSerTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	373
QY	1351	TTCTCTGTGATGGCTACGAGCTGACCCCGACAAGTGGACCTGCGAGCCCATCGAGCTG	1410
DB	374	PheLeuTrpMetGlyTyrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu	393
QY	1411	CCCGAGAGGAGAGCTGGACCGGTGAACGACATCCAGAAGCTGGTGGCGCAAGCTGAATGG	1470
DB	394	ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	413
QY	1471	GCAGCCAGATCTACCCCGCATCAAGGTGGCGGACGCTGTGCAAGCTGTGCGCGCGCC	1530
DB	414	AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr	433
QY	1531	AAGCCCTGACGACATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAGAAC	1590
DB	434	LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluGluGluAsn	453
QY	1591	CGCGAGATCTCGCGGAGCCGCGTGCACGCGTGTACTACGACCCCGAGAGACCTGGTG	1650
DB	454	ArgGluIleLeuArgIleProValHisGlyValTyrTyrAspProSerLysAspLeuVal	473
QY	1651	GCCGAGATCCAGAGAGGCGCCACGACCTGAGCTTACGAGATCTACAGGAGCCCTTC	1710
DB	474	AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe	493
QY	1711	AAGAACTGAGAGCCGCGAGTACCGCAGATCGGACCGCCCGACACCGACGCGTGAAG	1770
DB	494	LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg	513
QY	1771	CAGCTGACCGAGGCGCTGCAAGAGATCGCCATCGAGAGCATCTGTGCTGGGCAAGACC	1830
DB	514	GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr	533
QY	1831	CCCAAGTTCGCTGCCCATCCAGAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG	1890
DB	534	ProLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp	553
QY	1891	CAGGCCACTGGATCCCGAGTGGGAGTTCGTCAACACCCCGCCCTGGTGAAGCTCTGG	1950
DB	554	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp	573
QY	1951	TACAGCTGAGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC	2010
DB	574	TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla	593
QY	2011	AACCGGAGACCAAGATCGGCAAGCGCGCTAGTGACCGACCGGCGCGCGGAGAGATC	2070

Db 594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613  
Qy 2071 GTGAGCCTGACCGAGACACCAACAGACCGAGCTGCAGGCCATCCAGCTGCCCCTG 2130  
:::|||||  
Db 614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
:::|||||  
Qy 2131 CAGACAGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGATCATC 2190  
|||||  
Db 634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653  
|||||  
Qy 2191 CAGCCAGCCGACAGAGGAGGAGCTGTGTGAACACCATCATCGACGAGTCATC 2250  
|||||  
Db 654 GlnAlaGlnProAspArgSerGluSerGluValSerGlnIleIleGluGluLeuIle 673  
|||||  
Qy 2251 AAGAAGAGAGAGGTGTACCTGAGCTGGGTGCCGCCGCCCAAGGGCATCGCGGCAACGAG 2310  
|||||  
Db 674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693  
|||||  
Qy 2311 CAGATCGACAAGCTGGTGAGCAAGGGCATCGCGAAGGTGCTGCTCGGACGGCATCGAT 2370  
|||||  
Db 694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713  
|||||

## RESULT 13

US-10-296-734-1471  
; Sequence 1471, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A  
; TITLE OF INVENTION: Synthetic molecules and uses therefor  
; FILE REFERENCE: Savine  
; CURRENT APPLICATION NUMBER: US/10/296,734  
; CURRENT FILING DATE: 2003-08-04  
; PRIOR APPLICATION NUMBER: AU PQ7761/00  
; PRIOR FILING DATE: 2000-05-26  
; NUMBER OF SEQ ID NOS: 1507  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1471  
; LENGTH: 1006  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: POL Consensus A  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (15)..(15)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (22)..(23)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (26)..(26)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (30)..(30)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (36)..(36)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (38)..(39)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (41)..(44)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (311)..(333)  
; LOCATION: (46)..(46)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (48)..(48)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (51)..(51)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (53)..(55)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (58)..(58)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (73)..(73)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (75)..(75)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (105)..(105)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (130)..(130)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (173)..(173)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (194)..(194)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (197)..(197)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (255)..(255)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (280)..(281)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (293)..(293)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (300)..(300)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (316)..(316)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (320)..(320)  
; OTHER INFORMATION: X = unknown amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (331)..(333)

```
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (335)..(335)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (365)..(365)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (369)..(369)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (373)..(373)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(403)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (434)..(434)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (439)..(439)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (452)..(452)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (470)..(470)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (480)..(480)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (487)..(487)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (514)..(514)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (534)..(535)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (544)..(544)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (553)..(553)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (593)..(593)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (610)..(610)
; OTHER INFORMATION: X = unknown amino acid
```

```
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (675)..(675)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
```

```
Alignment Scores:
Pred. No.: 1,23e-152 Length: 1006
Score: 3289.50 Matches: 622
Percent Similarity: 89.36% Conservative: 25
Best Local Similarity: 85.91% Mismatches: 70
Query Match: 71.78% Indels: 2
DB: 15 Gaps: 2
```

US-09-610-313b-30 (11-2469) x US-10-296-734-1471 (11-1006)

```
QY 220 TTCTTCGGGAGGACCTGGCCTTCCCGGAGGAGGAGCCCGGAGTTCCTCCCGAGGACG 279
Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20
QY 280 AACCGCGCCAAACAGACCCACCGCGGAGCTGCAGTGGCGCGGACACACCC----- 333
Db 21 Thr*****AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp*****Leu 40
QY 334 CGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 378
Db 41 *****Gly***Glu***GlnGly***Gly*****SerPhe***PhePro 60
QY 379 CAGATCACCTGTGGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438
Db 61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80
QY 439 GCCTCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498
Db 81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLys 100
QY 499 TGAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 558
Db 101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGln 120
QY 559 ATCTGTATCGAGATCTCGCGCGAGAGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCG 618
Db 121 IleLeuIleGluIleCysGlyLysLys***IleGlyThrValLeuValGlyProThrPro 140
QY 619 GTCAACATCATCGCGCGCAACATGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 678
Db 141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160
QY 679 AGCCCATCGAGACCGTGCCTGGAAGTGAAGCCCGGCGATGAGCGCGCGCGCGCGCGCG 738
Db 161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180
QY 739 CAGTGGCCCTGACCGGAGGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798
Db 181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200
QY 799 AAGGAGGCGCAAGATCACCAAGATCGCGCGCGCGAGAACCCCTTACAACACCCCGTGTG 858
Db 201 LysGluGlyLysIleSerIysIleGlyProGluAsnProTyrrAsnThrProValPheAla 220
QY 859 ATCAAGAAGAGGACGACACCAAGTGGCGCGCAAGCTGGTGGAGCTTCGCGCGAGCTGA 918
Db 221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240
QY 919 CGCACCCGAGACTTCTGGGAGGTGCGAGCTGGGCGATCCCCACCGCGCGCGCGCGCG 978
Db 241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260
QY 979 AAGAAGAGCGTGCCTGCTGGACGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1038
Db 261 LysLysSerValThrValLeuAspValGlyAspAlaTyrrPheSerValProLeuAsp 280
QY 1039 GACTTCGCAAGTACACCGCCTTCCATCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1098
```



```
Db 281 ***PheArgLysTyrThrAlaPheThrIleProSer***AsnAsnGluThrProGly*** 300
Qy 1099 CGCTACCAAGTACAACGCTGCTCCCGCAGGCTGGAAGGCGACCCCGACGATCTTCCAGAGC 1158
Db 301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro***IlePheGln*** 320
Qy 1159 AGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACAG 1218
Db 321 SerMetThrLysIleLeuGluProPheArg*****Pro***IleValIleTyrGln 340
Qy 1219 TACATGACGACCTGTACGCTGGGAGCGACCTGGAGATCGCGCCAGCAGCCGCGCAAGATC 1278
Db 341 TyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360
Qy 1279 GAGGAGCTGCCAAGCAGCTCTCGCTGGCTGGGCTTCCACACCCCGCAAGAACGACCCAG 1338
Db 361 GluGluLeuArg***HisLeuLeu***TrpGlyPhe***ThrProAspLysLysHisGln 380
Qy 1339 AAGGAGCCCTTCTCTGTGGATCGGCTACGAGCTGCACCCCGCAAGAGTGACCGTCAG 1398
Db 381 LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln 400
Qy 1399 CCCATCAGCTGCCGAGAGGAGCTGGACCTGAACGATCCAGAGCTGGTGGC 1458
Db 401 Pro*****LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420
Qy 1459 AAGCTGAACCTGGGCGCAGCAGATCTACCCCGCATCAAGGTGCCCGCAGCTGTCAAGCTG 1518
Db 421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys***LysGlnLeuCys***Leu 440
Qy 1519 CTGCGCGCGCCCAAGGCTGACCGACATCGTCCCTGACCGAGGCGCGAGCTGGAG 1578
Db 441 LeuArgGlyAlaLysAlaLeuThrAspIleVal***LeuThrGluGluAlaGluLeuGlu 460
Qy 1579 CTGCGCGAGAACCGGAGATCTCTCGGAGCGCCCTGCACGCGCTGTACTAGACCCAGC 1638
Db 461 LeuAlaGluAsnArgGluIleLeuLys***ProValHisGlyValTyrTyrAspPro*** 480
Qy 1639 AAGCACTGTGTGGCGGAGATCCAGAAGCAGGCGCCACGACCATCGAGCTACCATCTAC 1698
Db 481 LysAspLeuValAlaGlu***GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500
Qy 1699 CAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATCGCGACCGCCACAC 1758
Db 501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla***LysArgSerAlaHisThr 520
Qy 1759 AACGACGTGAAGCAGCTGACCGGCGCTGCAGAAATCGCCATCGGAGCATCGTGATC 1818
Db 521 AsnAspValLysGlnLeuThrGluValValGlnLysVal*****GluSerIleValIle 540
Qy 1819 TGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1878
Db 541 TrpGlyLys***ProLysPheArgLeuProIleGln***GluThrTrpGlu***TrpTrp 560
Qy 1879 ACCGACTACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTG 1938
Db 561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580
Qy 1939 GTGAAGCTGTGGTACCAAGCTGGAAGAGGCCATCATCGCGCGGAGACCTTCTACGTG 1998
Db 581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle***GlyAlaGluThrPheTyrVal 600
Qy 1999 GACGCGCGCGCAACCGCGACCAAGATCGCAAGCGCGGCTACGTGACCGCGCGGGC 2058
Db 601 AspGlyAlaAlaAsnArgLysLys***GlyLysAlaGlyTyrValThrAspArgGly 620
Qy 2059 CGGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGAGAGCTGCAGCGCCATC 2118
Db 621 ArgGlnLysValValSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIle 640
Qy 2119 CAGCTGGCCCTGCAGGACGCGGAGGTGAACATCTGTGACCGCAGCGCATGACGCC 2178
```

```
Db 641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660
Qy 2179 CTGGGCTATCTCCAGGCGCCAGCCGACAAAGAGCGAGAGCTGGTGAACACGATCATC 2238
Db 661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu***ValAsnGlnIleIle 680
Qy 2239 GAGCAGCTGATCAAGAGGAGAGCTGTACTCTGAGCTGGGTGCCCGCCCAAGGGCATC 2298
Db 681 GluLysLeuIle***Lys***LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700
Qy 2299 GGCGGCAAGCAGCAGATCGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGTCTCTG 2358
Db 701 GlyGlyAsnGluGlnValAspLysLeuValSer***GlyIleArgLysValLeuPheLeu 720
Qy 2359 GACGCGCATCGAT 2370
Db 721 AspGlyIleAsp 724

RESULT 14
US-10-301-661A-6
; Sequence 6, Application US/10301661A
; Publication No. US20030157660A1
; GENERAL INFORMATION:
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-
; APPLICANT: INSERM
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS
; APPLICANT: INSTITUT PASTEUR
; APPLICANT: MAULIERE, Philippe
; APPLICANT: LOUSSERT-AJAKA, Ibtissem
; APPLICANT: SIMON, Francois
; APPLICANT: SARAGOSTI, Sentob
; APPLICANT: BARRE-SINOUSI, Françoise
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.
; FILE REFERENCE: 598US12
; CURRENT APPLICATION NUMBER: US/10/301.661A
; PRIOR FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US/09/319,588C
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: FR96/15087
; PRIOR FILING DATE: 1996-12-09
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PR1
; ORGANISM: Human immunodeficiency virus type 1
US-10-301-661A-6

Alignment Scores:
Pred. No.: 2,56e-149 Length: 1014
Score: 3221.50 Matches: 587
Percent Similarity: 90.80% Conservative: 74
Best Local Similarity: 80.63% Mismatches: 56
Query Match: 70.29% Indels: 11
DB: 14 Gaps: 3

US-09-610-313B-30 (1-2469) x US-10-301-661A-6 (1-1014)
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGGCCGCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgGluLeuTrpValSerGlyGluGlu 20
Qy 280 AAC-----CGCGCAACAGCCCGCCAGCGCGCGAGCTGCAGGTG----- 318
Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGluGlu 40
Qy 319 ---CGCGCGCAACACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60
Qy 367 CTGAACTTCCCGAGATCACCCTGTGGCAGCGCCCTGTGGATCATCAAGTGGCGCGCGCG 426
Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80
```



; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
 ; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
 ; FILE OF INVENTION: MODIFICATIONS  
 ; CURRENT APPLICATION NUMBER: US/09/952,060  
 ; CURRENT FILING DATE: 2001-09-14  
 ; PRIOR APPLICATION NUMBER: PCT/US01/28861  
 ; PRIOR FILING DATE: 2001-09-14  
 ; PRIOR APPLICATION NUMBER: 60/317,814  
 ; PRIOR FILING DATE: 2001-09-07  
 ; PRIOR APPLICATION NUMBER: 60/279,056  
 ; PRIOR FILING DATE: 2001-03-27  
 ; PRIOR APPLICATION NUMBER: 60/233,180  
 ; PRIOR FILING DATE: 2000-09-15  
 ; NUMBER OF SEQ ID NOS: 38  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 35  
 ; LENGTH: 1350  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Codon optimized gag-1A pol. fusion  
 US-09-952-060-35

Alignment Scores:  
 Pred. No.: 4.11e-145 Length: 1350  
 Score: 3135.50 Matches: 600  
 Percent Similarity: 81.29% Conservative: 43  
 Best Local Similarity: 75.85% Mismatches: 60  
 Query Match: 68.42% Indels: 88  
 DB: 10 Gaps: 5

US-09-610-313B-30 (1-2469) x US-09-952-060-35 (1-1350)

Qy 13 ATGCGCGAGCCATGAGCGAGCCACC---AGCGCCACATCTCTGATGCGAGCGAGCAAC 69  
 Db 363 LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArggLYsn 382  
 Qy 70 TTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGAGGAGGCGCACATCGCC 129  
 Db 383 PheArgAsnGlnArgLYsThrValLYsCysPheAsnCysGlyLYsValGlyHisIleAla 402  
 Qy 130 CGCAACTGCGCGCGCCCGCGCAAGAAGCGGTCTGTGAAGTGTGCGCAAGAGGCGCCACAG 189  
 Db 403 LysAsnCysArgAlaProArgLYsGlyCysTrpLYsCysGlyLYsGluGlyHisGln 422  
 Qy 190 ATGAAGGACTGCACCGAGCGCCAGCCCACTTCTTCCGCGAGGACCTGGCTTCCCGCCAG 249  
 Db 423 MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyLYsIleTrpProSerHisLY 442  
 Qy 250 GGCAGGCGCGGAGTTCCTCCAGCGAGCAGAAACCGCGCCCAACAGCCCGCCAGCGCGAG 309  
 Db 442 sGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProGluGluSe 462  
 Qy 310 CTGAGGTGCGCG-----GGGACAAACCCCGCGAGCGAGCGCGCGCGCGCGAGCGCGG 363  
 Db 462 rPheArgPheGlyGluGluLYsThrThrProSerGlnLYsGlnGluProIleAspLYs-- 481  
 Qy 364 ACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTCTGTGAGCATCAAGGTGGGC 423  
 Db 482 GluLeu-----TyroProLeuAlaSerLeuArg----- 490  
 Qy 424 GGCAGATCAAGGAGGCGCTGTGGACACCGCGCGCCAGACACCGTGTGGAGGAGATG 483  
 Db 490 ----- 490  
 Qy 484 AGCGTCCCGCAAGTGGAGGCCAAGATGATCGCGGCATCGCGCGCTTTCATCAAGGTG 543  
 Db 490 ----- 490  
 Qy 544 CGCCAGTACGACCCAGATCTGTGATGAGATCTGCGGCAAGAGGCCATCGGCGCGCTGTG 603  
 Db 490 ----- 490

Qy 604 ATCGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGTCACC 663  
 Db 491 -----SerLeuPheGlyAsnAspProSerSerGlnMetAla----- 502  
 Qy 664 CTGNACTTCCCATCAGCCCCATCAGACCGTGCCTGAGCTGAAGCTGAAGCCCGGATGAC 723  
 Db 503 -----ProfileSerProfileGluThrValProValLYsLeuLYsProGlyMetAsp 519  
 Qy 724 GGCCTCAAGGTGAAGTGGCGCTCAGCAGAGAGAGATCAAGGCGCTCAGCGCCATC 783  
 Db 520 GlyProLYsValLYsGlnTrpProLeuThrGluGluLYsIleLYsAlaLeuValGluIle 539  
 Qy 784 TGCAGGAGATGGAAGAGGCGCAAGATCACCAAGATCGCGCCCGGAAACCCCTACAAC 843  
 Db 540 CysThrGluMetGluLYsGluLYsIleSerLYsIleGlyProGluAsnProTyraen 559  
 Qy 844 ACCCGCGTTCGCCATCAAGAGAGAGAGACAGCAGCAGTGGCGCAAGCTGGTGGACTTC 903  
 Db 560 ThrProValPheAlaIleLYsLYsLYsAspSerThrLYsTrpArgLYsLeuValAspPhe 579  
 Qy 904 CGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCCCACC 963  
 Db 580 ArgGluLeuAsnLYsArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPro 599  
 Qy 964 GCGCGCTGAAGAGAGAGCGTGCCTGCTGACCGTGGCGCGAGCGCTACTTCCAGC 1023  
 Db 600 AlaGlyLeuLYsLYsLYsSerValThrValLeuAlaValGlyAspAlaTyPheSer 619  
 Qy 1024 GTGCGCTGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGCAGCATCAACAC 1083  
 Db 620 ValProLeuAspGluAspPheArgLYsThrAlaPheThrIleProSerIleAsnAsn 639  
 Qy 1084 GAGACCCCGCGCATCCGCTACAGTACAAAGTGTGCTGCCCGAGGCTGGAAGGCGAGCC 1143  
 Db 640 GluThrProGlyIleArgLYsGlnTrpAsnValLeuProGlnGlyTrpLYsGlySerPro 659  
 Qy 1144 AGCATCTTCCAGAGCAGCATGACCAAGATCCTGAGCGCTTCCGCGCGCCCAACCCGAG 1203  
 Db 660 AlaIlePheGlnSerSerMetThrLYsIleLeuGluProPheArgLYsGlnAsnProAsp 679  
 Qy 1204 ATCGTGTCTACCATGATGAGCGACCTGTACGTGGCGCGAGCTGGAGATCGGCGAG 1263  
 Db 680 IleValIleTrpGlnTrpMetAlaLeuValGlySerAspLeuGluIleGlyGln 699  
 Qy 1264 CACGCGCCCAAGATCGAGGAGCTGCGCAAGCAGCTGCTGCGCTGGCGCTTCCACACCC 1323  
 Db 700 HisArgThrLYsIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrPro 719  
 Qy 1324 GACAAGAGCAGCAGAGGAGCGCCCTTCTGTGGATGGCTACAGAGTGCACCCGAG 1383  
 Db 720 AspLYsLYsHisGlnLYsGluProProPheLeuTrpMetGlyTrpGluLeuHisProAsp 739  
 Qy 1384 AGTGGACCGTGCAGCGCCATCGAGCTGCCCGAGAGAGAGAGTGGAGCGTGAACAGCATC 1443  
 Db 740 LysTrpThrValGlnProfileValLeuProGluLYsAspSerTrpThrValAsnAspIle 759  
 Qy 1444 CAGAAGCTGTGGCGAAGTGAATGGCGCAGCAGATCTACCCCGCGCATCAAGGTGGCG 1503  
 Db 760 GlnLYsLeuValGlyLYsLeuAsnTrpAlaSerGlnIleTrpProGlyIleLYsValArg 779  
 Qy 1504 CAGCTGTGCAAGTGTGCGCGCGCCAGCGCTGACCGCATCTGCGCTGCCCTGACCGAG 1563  
 Db 780 GlnLeuCysLYsLeuLeuArgGlyThrLYsAlaLeuThrGluValIleProLeuThrGlu 799  
 Qy 1564 GAGCGAGCTGAGCTGGCGCGAGAACCGCGAGATCTCGCGCGAGCGCGCTGACCGGCTG 1623  
 Db 800 GluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLYsGluProValHisGlyVal 819  
 Qy 1624 TACTACGACCCCGAGGAGGAGCTGGTGGCGGAGATCCAGAGGAGGCGCCAGCAGCTGG 1683  
 Db 820 TyrtAspProSerLYsAspLeuIleAlaGluIleGlnLYsGlnGlyGlnGlyGlnTrp 839

QY 1684 ACCTACCAGATCTACCAGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCAAGATG 1743  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 1744 CGCACCGCCACACCAACGAGCTGAAGACGCTGACCGAGGCCCTGCGAAGATCGCCATG 1803  
 Db |||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 1804 GAGAGCATCTGATCTGGGCGACAGACCCCAAGTTCGGCTGCCATCCAGAGGAGGACC 1863  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 1864 TGGGAGACCTGGTGGACCGGCTACTGCGACGCCACCTGGATCCCGAGTGGGAGTTCGTG 1923  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 1924 AACACCCCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGGGCC 1983  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 1984 GAGACCTTCTACGTGACGCGCGCGCCCAACCGGAGACCAAGATCGCGAGGCGGCTAC 2043  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 2044 GTGACCGACCGGCGCGGAGAGATCGTGCAGCTGACCGAGACCAACCAAGAGGACC 2103  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 2104 GAGCTGAGGCGCCATCCAGCTGGCCCTGCGAGCAGCGGCGAGGAGTGAACATCGTGACC 2163  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 2164 GAGACCGAGTACGCGCTGGGCGATCATCCAGCGCGCGCGCGAGAGGAGGAGGAGCTG 2223  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 2224 GTGAACCGAGATCATCGAGCAGCTGATCAAGAAGGAGAGGAGTGTACCTGAGCTGGTGCCC 2283  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 2284 GCCCACAGGCGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGCATCCGC 2343  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 2344 AAGTGCTGTCTCTGGACGCGCATCGAT 2370  
 Db |||||||TyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 2370 LysValLeuPheLeuAspGlyIleAsp 1068

Search completed: June 2, 2005, 05:27:02  
 Job time : 245.54 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 20:20:55 ; Search time 209.175 Seconds  
(without alignments)  
12088.668 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 4583

Sequence: 1 gtcagccaccatggcga.....gggctagcaccggtgaattc 2469

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2 1/USPTO.epool/US09610313/runat 31052005 155136 15147/app query.fasta.1.7893  
-DB-UniProt\_03 -QFWT-fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US09610313 @CGN 1 1 725 @runat 31052005 155136 15147 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOQUY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4178	91.2	1427	2 Q9WF92	Q9wf92 human immun
2	4174	91.1	1427	2 Q9WF89	Q9wf89 human immun
3	4170	91.0	1428	2 Q9WF86	Q9wf86 human immun
4	4028	87.9	1427	2 Q9WF62	Q9wf62 human immun
5	4021	87.7	1429	2 Q9X4F8	Q9x4f8 human immun
6	3993	87.1	1425	2 Q9X4F6	Q9x4f6 human immun
7	3986	87.0	1427	2 Q9WF96	Q9wf96 human immun
8	3982	86.9	1430	2 Q9X4R4	Q9x4r4 human immun
9	3973.5	86.7	1427	2 Q9WF71	Q9wf71 human immun
10	3969	86.6	1427	2 Q9WF60	Q9wf60 human immun
11	3966.5	86.5	1427	2 Q9WF65	Q9wf65 human immun
12	3965	86.5	1437	2 Q9WF77	Q9wf77 human immun
13	3957.5	86.4	1427	2 Q9WF74	Q9wf74 human immun
14	3951.5	86.2	1427	2 Q9ADX5	Q9adx5 human immun
15	3943	86.0	1425	2 Q9WF79	Q9wf79 human immun
16	3936	85.9	1427	2 Q9WF98	Q9wf98 human immun

#### ALIGNMENTS

RESULT 1

Q9WF92	17	3931.5	85.8	1429	2	Q9WF84	Q9wf84 human immun
Q9WF92	18	3922	85.6	1428	2	Q9X4P0	Q9x4p0 human immun
Q9WF92	19	3912.5	85.4	1427	2	Q9WF68	Q9wf68 human immun
Q9WF92	20	3876.5	84.6	1432	2	Q9ADZ9	Q9adz9 human immun
Q9WF92	21	3873	84.5	1433	2	Q9AD89	Q9ad89 human immun
Q9WF92	22	3871	84.5	1434	2	Q9ADV1	Q9adv1 human immun
Q9WF92	23	3869	84.4	1430	2	Q7SV36	Q7sv36 human immun
Q9WF92	24	3865	84.3	1425	2	Q7SV20	Q7sv20 human immun
Q9WF92	25	3859.5	84.2	1433	2	Q9ADN7	Q9adn7 human immun
Q9WF92	26	3858.5	84.2	1433	2	Q9ADK5	Q9adk5 human immun
Q9WF92	27	3856.5	84.1	1433	2	Q9AD87	Q9ad87 human immun
Q9WF92	28	3853.5	84.1	1433	2	Q9ADG5	Q9adg5 human immun
Q9WF92	29	3850.5	84.0	1435	2	Q9DQ33	Q9dq33 human immun
Q9WF92	30	3849.5	84.0	1433	2	Q9ADE1	Q9adel1 human immun
Q9WF92	31	3844.5	83.9	1431	2	Q9ADR1	Q9adr1 human immun
Q9WF92	32	3843.5	83.9	1433	2	Q9ADCS	Q9adc5 human immun
Q9WF92	33	3842.5	83.8	1433	2	Q9AE32	Q9ae32 human immun
Q9WF92	34	3841.5	83.8	1433	2	Q9ADM1	Q9adml1 human immun
Q9WF92	35	3841.5	83.8	1433	2	Q9ADU3	Q9adu3 human immun
Q9WF92	36	3840.5	83.8	1437	2	Q41782	Q41782 human immun
Q9WF92	37	3839	83.8	1434	2	Q9WLJ3	Q9wlj3 human immun
Q9WF92	38	3839	83.8	1435	2	Q9WLJ3	Q9wlj3 human immun
Q9WF92	39	3838.5	83.8	1433	2	Q9AD85	Q9ad85 human immun
Q9WF92	40	3836.5	83.7	1433	2	Q7SPT9	Q7sept9 human immun
Q9WF92	41	3832.5	83.6	1433	2	Q9ADW7	Q9adw7 human immun
Q9WF92	42	3832.5	83.6	1436	2	Q9AD93	Q9ad93 human immun
Q9WF92	43	3830.5	83.6	1433	2	Q9AE23	Q9ae23 human immun
Q9WF92	44	3829.5	83.6	1433	2	Q9ADR9	Q9adr9 human immun
Q9WF92	45	3827.5	83.5	1433	2	Q9ADH3	Q9adh3 human immun

Q9WF92 PRELIMINARY; PRT; 1427 AA.  
AC Q9WF92; PRT; 1427 AA.  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OC NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahma M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana";  
RT J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110975; AAD17135.1; -;  
DR HSSP; P04585; 1JLA.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.



Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
 Qy 1573 CTGGAGCTGGCGGAGAAACCGGAGATCTCTGGCGAGCCCGTGCACCGCGTGTACTACGAC 1632  
 Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrAsp 899  
 Qy 1633 CCCAGCAGGACCTGGTGGCGGAGATCCAGAGCAGGCGCCAGCCAGCTGACCTACACG 1692  
 Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
 Qy 1693 ATCTACACGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCAAGATGCGCACCCGCC 1752  
 Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
 Qy 1753 CACACCAACGAGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATCGAGAGCATC 1812  
 Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
 Qy 1813 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGAGACC 1872  
 Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
 Qy 1873 TGGTGGACCGACTACTGCGAGCCACCTCGGATCCCGAGTGGAGTTCGTGAACACACCCCC 1932  
 Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
 Qy 1933 CCCCTGTGTGAAGTGTGTACAGCTGAGAGGAGCCCATCATCGGCGCGAGACCTTC 1992  
 Db 1000 ProLeuValLysLeuTyrTrpGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
 Qy 1993 TACGTGACGCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTACTGTCACCGAC 2052  
 Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
 Qy 2053 CGGCGCGCGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACGAGCTGCAG 2112  
 Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrAsnGlnLysThrGluLeuGln 1059  
 Qy 2113 GCCATCCAGCTGGCGCTCGAGACAGCGCGGAGGAGTGAACATCGTGACCGACGACG 2172  
 Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
 Qy 2173 TACGCGCTGGCATCATCAGCCCGCCAGCCGCGACAGAGCGAGCGAGCTGTGACACG 2232  
 Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
 Qy 2233 ATCATCGAGCAGCTGATCAAGAAGGAGAGTGTACCTGAGCTGGTCCCGCCACACAG 2292  
 Db 1100 IleIleGlnLeuIleLysGlnLysValTyrLeuSerTrpValProAlaHisLys 1119  
 Qy 2293 GGCATCGCGCGCAACGAGCAGATCGACAGCTGTGAGCAAGGGCATCCGCAAGGTGCTG 2352  
 Db 1120 GlyIleGlyLysAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139  
 Qy 2353 TTCCTGACGCGCATCGAT 2370  
 Db 1140 PheLeuAspGlyIleAsp 1145

## RESULT 2

Q9WF89 PRELIMINARY; PRT; 1427 AA.

AC Q9WF89;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gas-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99214383; PubMed=10196340;

RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,

RA Folley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.,  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana.",  
 RL J. Virol. 73:4427-4432(1999).  
 CC -I- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110974; AAD17126.1; -;  
 DR HSP; P04585; 1JLA.  
 DR GO: GO:0019012; C:virion; IEA.  
 DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR GO: GO:0008907; F:integrase activity; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO: GO:0003723; F:RNA binding; IEA.  
 DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR GO: GO:0016740; F:transferase activity; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0015074; P:DNA integration; IEA.  
 DR GO: GO:0006310; P:DNA recombination; IEA.  
 DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO: GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_P24.  
 DR InterPro; IPR001037; Integrase C.  
 DR InterPro; IPR003308; Integrase Zn N.  
 DR InterPro; IPR001995; Peptidase A2.  
 DR InterPro; IPR009007; Pept. Aspartic.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR010999; Retrovir matrix.  
 DR InterPro; IPR000071; Retrovir\_p17.  
 DR InterPro; IPR008916; Retrov capsid C.  
 DR InterPro; IPR008919; Retrov capsid N.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVtse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_P24; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIV1MATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS0158; ZF\_CCHC; 2.  
 KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.

## Alignment Scores:

Pred. No.: 3,87e-159 Length: 1427  
 Score: 4174.00 Matches: 784  
 Percent Similarity: 99.87% Conservative: 1  
 Best Local Similarity: 99.75% Mismatches: 1  
 Query Match: 91.08% Indels: 0  
 DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x Q9WF89 (1-1427)

Qy 13 ATGCCGAGGCGCATGAGCCAGCCAGCCCAACATCTCTGTATGTCAGCGCACTTC 72



Db 360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379  
:::|||||  
Qy 73 AAGGGCCCCAAGCGCATCAAGTCTTCAACTGGCGCAAGGAGGGCCACATCGCCGC 132  
Db 380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
Qy 133 AACTGCGCGCCCCCAAGAGGGCTGCTGGAAGTGGCGCAAGGAGGGCCACCAAGATG 192  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
Qy 193 AAGGACTGCACCGAGCGCCAGGCCAACTTCTTCGGAGAGCACTTGCGCTTCCCCAGGGC 252  
Db 420 LysAspCysThrGluArgGlnAlaAsnPheArgGluAspLeuAlaPheProGlnGly 439  
Qy 253 AAGGCGCGCAGTCTCCACGAGAGCAGAACCGCGCCACAGCCCCACCAGCGCGAGCTG 312  
Db 440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459  
Qy 313 CAGGTGGCGCGCACAAACCCCGCAGCGAGCGCGCGCCGAGCGCCAGGGCACCTGAAC 372  
Db 460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479  
Qy 373 TTCCCCCAGATCACTCTGTGGCAGCGCCCTCTGGTAGCATCAAGGTGGCGCCAGATC 432  
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499  
Qy 433 AAGGAGCCCTGTGGACACCGCGCGCGCAGCACACCGTGTGGAGGAGATGAGCTGCC 492  
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519  
Qy 493 GGCAGGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCAATCAAGGTGGCCAGTAC 552  
Db 520 GlyLysTrpLysProLysMetIleGlyIleGlyPheIleLysValArgGlnTyr 539  
Qy 553 GACCAGATCTGTATCGAGATCTCGGCAAGAAGGCCATCGGCACCGTGTGTATCGGCC 612  
Db 540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559  
Qy 613 ACCCGGTGAACATCATCGCGCGCGCAATCTGTCACCCAGTGGCTGGCTGACCTTTC 672  
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579  
Qy 673 CCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAG 732  
Db 580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy 733 GTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTCGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
Qy 793 ATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
Qy 853 TTCGCCATCAAGAAGAGGACAGCACCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAAAGCGCACCCAGGACTTCTGGGAGGTGAGCTGGGCATCCCCACCCCGCGGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAGAGAGCTGACGTGCTGGAGCTGGGAGACCGCTACTTCAGCGTGGCCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGACTTCGCGAGTACACCGCTTCACCATCCCGAGCATCAACACAGAGACCCC 1092  
Db 700 AspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnGluThrPro 719  
Qy 1093 GGCATCGCTACCAAGTACAACTGCTGCCCGCGGCTGGAGGGCAGCCCCAGCATCTTC 1152  
|||||

Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
Qy 1213 TACCACTACATGGACGACCTGTACGTGGGACGACCTCGAGATCGCCAGCACCGCGCC 1272  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1273 AGATCGAGAGCTGCGCAAGCACCTCTGCTGGCTGGGGTTCACCAACCCCGCAGAGAAG 1332  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
Qy 1333 CACCAAGAGGAGCGCCCTTCTGTGTGGATGGGTACGAGCTGCACCCCGACAGTGGACC 1392  
Db 800 HisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1393 GTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGACCGTGAACGACATCCAGAAGCTG 1452  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1453 GTGGGCAAGCTGAACCTGGGCGCAGCATCTACCCCGGCATCAAGGTGGCGCAGCTGTGC 1512  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuCys 859  
Qy 1513 AAGCTCTGGCGCGCCCAAGGCCCTGACCGCATCGTGCCTGACCGAGGAGGCGGAG 1572  
Db 860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
Qy 1573 CTGGAGCTGGCGGAGAACCGCGAGATCTTGGCGGAGCCCGTGCCAGCGCTGTACTACGAC 1632  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899  
Qy 1633 CCAGCAAGAGACTGTGTGGCGGAGATCCAGAGCAGGGGCCACACAGATGGAGCTACCAG 1692  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1693 ATCTACCAAGGAGCCCTTCAAGACCTGAAGCCGCAAGTACGCCAAGATGCGCACCGCC 1752  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1753 CACCAACACGATGAAGCAGCTGACCGAGGCGGTGACAGAGATCGCCATCGGAGAGCATC 1812  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
Qy 1813 GTGATCTGGGCGCAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGAGCTGGGAGACC 1872  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1873 TGTGTGACCGACTACTGCGAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1932  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1933 CCCTCGTGAAGCTGTGTACCAAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACTTC 1992  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
Qy 1993 TAGCTGGAGCGCGCCCGCAACCGCGGAGACCAAGATCGGCAAGCGCCGGCTACGTGACCGAC 2052  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2053 CGGGCGCGCAGAAAGATCTGTAGCTGACCGAGACCAACCAACAGAAAGACCGAGCTGCAG 2112  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
Qy 2113 GCCATCCAGCTGGCCCTGCGAGGACAGCGCGCAGAGGTGAACATCGTGACCGCAGCCAG 2172  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2173 TAGCCCTGGCGATCATCCAGGCCCGCGCAGCAAGAGCGAGAGCGAGCTGGTGAACCCAG 2232  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluLeuValAsnGln 1099  
|||||

Qy 2233 ATCATCGAGCGCTGATCAAGAGAGAGAGTGTACTGAGCTGGTCCCGCCACCAAG 2292  
 Db 1100 lfeileGluInLeulleLysGluLysValtyrLeuSerTrpValProAlaHisLys 1119  
 Qy 2293 GGCATCGCGCGCAGCAGCAGTGTGAGCAAGGCGATCCGCAAGTGCTG 2352  
 Db 1120 GlyileGlyLysGluInleAspLysLeuValSerLysGlyileArgLysValLeu 1139  
 Qy 2353 TTCTCGAGCGCATCGAT 2370  
 Db 1140 PheLeuaspGlyileAsp 1145  
 RESULT 3  
 Q9WF86  
 ID Q9WF86 PRELIMINARY; PRT; 1428 AA.  
 AC Q9WF86;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana."  
 RL J. Virol. 73:4427-4432(1999).  
 CC -I- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110973; AAD17119.1; -.  
 DR HSP; P04585; 1JLA.  
 DR GO; GO:0019012; C:virion; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:integrase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0015074; P:DNA integration; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn\_N.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR009007; Pept\_Aspatic.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR010999; Retrovir\_matrix.  
 DR InterPro; IPR000071; Retrovir\_p17.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR004477; RVTse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 Pfam; PF00540; Gag\_p17; 1.  
 Pfam; PF00607; Gag\_p24; 1.  
 Pfam; PF02022; Integrase\_Zn; 1.  
 Pfam; PF00075; RNaseH; 1.

Pfam; PF00665; rve; 1.  
 Pfam; PF00077; RVP; 1.  
 Pfam; PF00078; RVT\_1; 1.  
 Pfam; PF06815; RVT\_connect; 1.  
 Pfam; PF06817; RVT\_thumb; 1.  
 Pfam; PF00098; zf\_CCHC; 2.  
 PRINTS; PR00399; C2HCZNFINGER.  
 PRINTS; PR00234; HIVIMATRIX.  
 SMART; SM00343; Znf\_C2HC; 2.  
 PROSITE; PS00141; ASP\_PROTASE; 1.  
 PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
 PROSITE; PS0158; ZF\_CCHC; 2.  
 AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 1428 AA; 161328 MW; 2D61CD9FEABD2387 CRC64;  
 Alignment Scores:  
 Pred. No.: 559e-159 Length: 1428  
 Score: 4170.00 Matches: 783  
 Percent Similarity: 100.00% Conservative: 3  
 Best Local Similarity: 99.62% Mismatches: 0  
 Query Match: 90.99% Indels: 0  
 DB: 2 Gaps: 0  
 US-09-610-313B-30 (1-2469) x Q9WF86 (1-1428)  
 Qy 13 ATGCGCGAGCGCATGAGCGCAGCCACAGCCCAACATCTCTGATGCGAGCAACTTC 72  
 Db 360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAlaLeuMetGlnArgSerAsnPhe 379  
 Qy 73 AAGCGCCCCAAGCCATCATCACTGCTTCACTGCGCAAGAGGGCCACATCGCCGC 132  
 Db 380 LysGlyProLysArgIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
 Qy 133 AACTGCGCGCGCGCCGAGAGGGTCTGGAAGTGGCGCAGAGGGCCACCAGATG 192  
 Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
 Qy 193 AAGGACTGCACCGAGCGCGCCCACTTCTCCGCGAGGACCTGGCTTCCCCCAGAGC 252  
 Db 420 LysAspCysThrGluArgGlnAlaAsnPheArgGluAspLeuAlaPheProGlnGly 439  
 Qy 253 AAGCGCGCGAGTTTCCCGAGCAGCAGAACCGCCCAACAGCCCCCAGCGCGAGCTG 312  
 Db 440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459  
 Qy 313 CAGTGGCGCGGACAAACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372  
 Db 460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479  
 Qy 373 TTCCCGCAGATCACCTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 432  
 Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499  
 Qy 433 AAGGAGCGCGCTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 492  
 Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519  
 Qy 493 GGCNAGTGGAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552  
 Db 520 GlyLysTrpLysProLysMetMetGlyGlyIleGlyPheIleLysValArgGlnTyr 539  
 Qy 553 GACCAGATCTGTATCGAGATCTGGCGCAAGAGGCCCATCGCACCGTGTCTCATCGGCC 612  
 Db 540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559  
 Qy 613 ACCCGCGTGAACATCATCGCGCGCGCAACATGTCACCCAGCTGGGCTGCACCTGAAC 672  
 Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579  
 Qy 673 CCNATGAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCGCGCGCGCGCGCGCGCG 732

Db 580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
QY 733 GTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGCCCTGACCGCATCTCGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysLeuAlaLeuThrAlaIleCysGluGlu 619  
QY 793 ATGAGAAAGAGGAGCAAGATCACCAAGATCGGCCCGCCGAGAACCCCTACAAACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
QY 853 TTCGCCATCAAGAAGAAAGACAGACCAAGTGGCGCAAGCTGTGGATCTTCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
QY 913 AACAAAGCGCACCGAGACTTCTGGAGGTGACGTGGGCATCCCGACCCCGCGCCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
QY 973 AAGAAGAAGAGCGTGCCTGCTGGAGCTGGCGACCGCTACTTTCAGCGTGCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
QY 1033 GACGAGACTTCCGCAAGTACACCGCTTCACCATCCCCAGCATCAACAAAGACACCCC 1092  
Db 700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GGCATCGCTACCAAGTACAGCTGCTGCCCGCTGGAGGCGACCGCCAGCATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
QY 1153 CAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCCCGCAACCCCGAGATCTGTATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
QY 1213 TACAGTACATGACGACCTGTGTAGTGGGACGACCTGAGATCGCGCAGCACCGGCC 1272  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
QY 1273 AAGATCGAGGAGCTGCGCAGCACCTGCTGGCTGGGGCTTACCACCCCGCAGCAAG 1332  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrProAspLysLys 799  
QY 1333 CACCAGAAAGAGCCCCCTTCTGTGTGATGGGTACGAGCTGCACCCCGCACAAGTGAC 1392  
Db 800 HisGlnLysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTrpThr 819  
QY 1393 GTGACGCCCATCGAGCTGCCCGCAGAGGAGGAGTGTGAACCTGACATCCAGAGCTG 1452  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
QY 1453 GTGGCAAGCTGAACCTGGCGCAGCCAGATCTACCCCGGCATCAAGTGGCCAGCTGTGC 1512  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLys 859  
QY 1513 AAGTGTGTGGCGCGCCCAAGGCCCTGACCGACATCTGCGCCCTGACCGAGAGGCCGAG 1572  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1573 CTGGAGCTGCCGAGAACCGCAGATCTCTGGGAGCCCTGACCGCGGTGTACTAGAC 1632  
Db 880 LeuGluLeuAlaGluAspArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899  
QY 1633 CCCAGCAGGACCTGTGGCGGAGATCCAGACGAGGGCCACACCATGTGACCTACCAG 1692  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
QY 1693 ATCTACGAGGAGCCCTTCAAGAACCTCAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1752  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
QY 1753 CACACCAACAGCTGTGAAGCAGCTGACCGAGCCGTGCAGAAAGATCGCCATGTGAGAGATC 1812  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

QY 1813 GTGATCTGGGGCAAGACCCCAAGTTCCGCTCCCATCCAGAAAGAGACCTGGGAGCC 1872  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1873 TGGTGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1932  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY 1933 CCCTCGTGAAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGCGCGGAGACCTTC 1992  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
QY 1993 TAGCTGCAGCGCGCCCAACCGCAGAGCAACATCGCAAGCCCGCTACGTGACCCGAC 2052  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
QY 2053 CGGGCGCGCAGAGATCTGTGACCTGACCGAGACCAACCAAGACCGAGCTGCAG 2112  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
QY 2113 GCATTCAGCTGGCCCTGCGAGGACGCGGAGGTGAACATCGTACCGCAGACGAC 2172  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
QY 2173 TAGCCCTGGGCATCATCCAGGCCCGCAGCCGACAGAGCGAGCGAGCTGGTGAACCCAG 2232  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
QY 2233 ATCATCGAGCAGCTGTATCAAGAGGAGAGTGTACTGAGCTGGTGGTCCCGCCCAAG 2292  
Db 1100 IleIleGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2293 GGCATCGCGCGCAACGACGAGATCGCAAGCTGGTGAAGGCGCATCCGCAAGTGTCTG 2352  
Db 1120 GlyIleGlyLysGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139  
QY 2353 TTCCTCGAGCGCATCGAT 2370  
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 4  
Q9WF62 PRELIMINARY; PRT; 1427 AA.  
ID Q9WF62 AC Q9WF62; DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.,  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana."; J. Virol. 73:4427-4432(1999).  
RL CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110961; AAD17021.1; -.  
DR HSSP; Q70622; IHVN.  
DR DR GO:0019012; C:virion; IEA.  
DR DR GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR DR GO:0003677; F:DNA binding; IEA.  
DR DR GO:0008907; F:integrase activity; IEA.  
DR DR GO:0008233; F:peptidase activity; IEA.  
DR DR GO:0004523; F:ribonuclease H activity; IEA.  
DR DR GO:0003723; F:RNA binding; IEA.  
DR DR GO:0003964; F:RNA-directed DNA polymerase activity; IEA.



```
QY 1393 GTGAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGACTG 1452
Db 820 ValGlnProIleuAenLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839
QY 1453 GTGGCAAGCTGAACCTGGGCCAGCAGCATATACCCCGGCATCAAGGTGCGCCAGCTGTGC 1512
Db 840 ValGlyLysLeuAenTrpAlaSerGlnIleTyProGlyIleLysValArgGlnLeuCys 859
QY 1513 AAGCTGTCGGCGGCCCAAGGCCCTGACCGACATCGTGCCCTGACGAGAGGCGCGAG 1572
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879
QY 1573 CTGGAGCTGCCGAGAACCGGAGATCTCGGCGAGCCCTGACCGCGTGTACTAGCAC 1632
Db 880 LeuGluLeuAlaGluAenArgGluIleLeuLysGlnProValHisGlyValTyTrpAsp 899
QY 1633 CCAGCAAGACCTGTGTGGCCGAGATCCAGAAGCAGGCCACACGACGAGCTACCCAG 1692
Db 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrgln 919
QY 1693 ATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC 1752
Db 920 IleTyrglnGluProPheLysAsnLeuLysThrGlyLysTyrglyLysMetArgThrAla 939
QY 1753 CACACCAACGACGTGAGCAGCTGACCGAGCGCGCTGCAGAAGATCGCCATGGAGAGCATC 1812
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959
QY 1813 GTGATCTGGGGCAAGACCCCAAGTTCGCGCTCGCCATCCAGAGGAGACCTGGGAGACC 1872
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979
QY 1873 TGTGAGACCACTACTGCGCAGGCCACTGTGATCCCGAGTGGAGTTCGTGAACACCCCC 1932
Db 980 TrpTrpThrAspTyTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999
QY 1933 CCCTGTGTGAAGCTGTGTACCAAGCTGGAGAGGAGCGCCATCATCGCGCGGAGACCTTC 1992
Db 1000 ProLeuValLysLeuTrpTyrglnLeuGluGluGluProIleValGlyAlaGluThrPhe 1019
QY 1993 TAGCTGGACCGCGCGCCAAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2052
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrglyValThrAsp 1039
QY 2053 CGGGCGCGGAGAGATCGTGAAGCTGACCGAGACACCAACAGAGACCGAGCTGCAG 2112
Db 1040 ArgGlyArgGlnArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059
QY 2113 GCCATCCAGCTGCGCTGCAGGACAGCGGAGCGAGTGAACATCGTGACCGACGACGAG 2172
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079
QY 2173 TAGCGCTGGGCGCATCTCCAGGCGCCAGCCGACAAAGAGCGAGCGAGCTGGTGAACGAG 2232
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099
QY 2233 ATCATCGAGAGCTGATCAAGAGGAGAGAGGTGATCTGAGCTGGGTGCGCCGCGCCAGAG 2292
Db 1100 IleIleGluGlnLeuIleLysLysGluArgValTyrglySerTrpValProAlaHisLys 1119
QY 2293 GGCATCGGGCGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCCGCAAGGTGCTG 2352
Db 1120 GlyIleGlyLysGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
QY 2353 TTCCTGGACGGCATCGAT 2370
Db 1140 PheLeuAspGlyIleAsp 1145
RESULT 5
Q6X4P8 PRELIMINARY; PRT; 1429 AA.
ID Q6X4P8
AC Q6X4P8;
```

```
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Gag-poi fusion polyprotein (Fragment).
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14709249; DOI=10.1089/0892220371881220;
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birk D.L.,
RA McCutchan F.E.;
RT "A Cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT in full genome analysis, is not sustained in subgenomic regions.";
RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Harris M.E., Birk D.L.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim B., McCutchan F.E.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Maayan S., Zeira M.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to peptidase family A2.
DR HSPB; P03367; 1A8G.
DR GO: GO:0019012; C:virion; IEA.
DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0008907; F:peptidase activity; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004523; F:ribonuclease H activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0015074; P:DNA integration; IEA.
DR GO: GO:0006310; P:DNA recombination; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO: GO:0016032; P:vital life cycle; IEA.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR001037; Integrase_C.
DR InterPro: IPR003308; Integrase_Zn_N.
DR InterPro: IPR001995; Peptidase_A2.
DR InterPro: IPR009007; Pept_Aspatic.
DR InterPro: IPR001969; Pept_Asp_AS.
DR InterPro: IPR010999; Retrovir_matrix.
DR InterPro: IPR000071; Retrovir_p17.
DR InterPro: IPR008916; Retrov_capsid_C.
DR InterPro: IPR002156; Retrov_capsid_N.
DR InterPro: IPR001584; RNaseH.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR010659; RVT_connect.
DR InterPro: IPR010661; RVT_thumb.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00540; Gag_p17; 1.
DR Pfam: PF02022; Integrase_Zn; 1.
DR Pfam: PF00075; RNaseH; 1.
DR Pfam: PF00665; rve; 1.
DR Pfam: PF00077; RVP; 1.
DR Pfam: PF00078; RVT_1; 1.
DR Pfam: PF06815; RVT_connect; 1.
DR Pfam: PF06817; RVT_thumb; 1.
DR Pfam: PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
```

DR PRINTS; PR00234; HIV1MATRIX.  
 DR SMART; SM00343; ZnF\_C2HC; 2.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS0175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS0158; ZF\_C2HC; 2.  
 KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; transferase; Zinc;  
 KW Zinc-finger.  
 FT NON TER 1  
 SQ SEQUENCE 1429 AA; 161586 MW; DOB9CD124DC64E19 CRC64;

Alignment Scores:  
 Pred. No.: 5,02e-153 Length: 1429  
 Score: 4021.00 Matches: 747  
 Percent Similarity: 97.97% Conservative: 27  
 Best Local Similarity: 94.56% Mismatches: 12  
 Query Match: 87.74% Indels: 4  
 DB: 2 Gaps: 1

US-09-610-313B-30 (1-2469) x Q6X4P8 (1-1429)

Qy	13	ATGCCGAGGCGCATCAGCAGCCAGCCAGCCCAACATCTCTGATCGACGCGCAGCAACTTC	72
Db	358	LeuAlaGluAlaMetSerGlnAlaAsnAlaHisMetMetGlnLysSerAsnPhe	377
Qy	73	AAGGGCCCAAGCGCATCATCAAGTGTCTCAACTGCGGCAAGAGGGCCACATCGCCGC	132
Db	378	LysGlyProLysArgIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg	397
Qy	133	AACCTGCGCGCCCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCATG	192
Db	398	AsnCysArgAlaProArgLysArgGlyCysTrpLysCysGlyLysGluGlyHisGlnMet	417
Qy	193	AAGGACTGCACCGAGCGCAGCCCAACTCTTCGCGCAGGACCTCGGCCTTCCGCCAGGGC	252
Db	418	LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGly	437
Qy	253	AAGCGCGCGAGTTCGCCAGCAGAGAACCGCGCCCAACAGCCCAACAGCGCGAGCTG	312
Db	438	GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu	457
Qy	313	CAGGTGCGCGGACACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	363
Db	458	GlnValArgGlyAspAsnProIleSerGluAlaGlyAspGluArgGlnGlyAlaLeuGln	477
Qy	364	---ACCTGAACTTCCCGCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTG	420
Db	478	ArgThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle	497
Qy	421	GGCGCCAGATCAAGGAGGCGCTCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCG	480
Db	498	GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu	517
Qy	481	ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCATCGCGGCTTCATCAAG	540
Db	518	IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyPheIleLys	537
Qy	541	GTGCGCAGTACGACCATCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCACCGTG	600
Db	538	ValArgGlnTrpAspGlnIleLeuIleGluLysCysGlyLysLysAlaIleGlyThrVal	557
Qy	601	CTGATCGCGCCCAACCCCGTGAAACATCATCGCGCGCAACATGCTGACCCAGCTGGGTGC	660
Db	558	LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys	577
Qy	661	ACCTGAACTTCCCATCAGCCCATCGAGACCTGCGCGTGAAGCTGAAGCCCGCGCATG	720
Db	578	ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet	597
Qy	721	GACGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGGCCCTCACCGCC	780
Db	598	AspGlyProLysValLysGlnTrpProLeuThrLysGluLysIleGluAlaLeuThrAla	617

Qy	781	ATCTGCGAGGATGGAGAGGAGGCGCAGATCACCAAGATCGGCCCCGAGAACCCCTAC	840
Db	618	IleCysGluGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyr	637
Qy	841	AACACCCCGTGTTCGCATCAAGAAGACAGACCAAGTGGCGCAAGCTGTGTGAC	900
Db	638	AsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAsp	657
Qy	901	TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGGGAGGTGCAGCTGGGCATCCCCAC	960
Db	658	PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis	677
Qy	961	CCCCCGCCCTGAAGAAGAAGAGCGTGCAGCTGTGGAGCTGGGCGCGCTACTTTC	1020
Db	678	ProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe	697
Qy	1021	AGCGTGGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCCAGCATCAAC	1080
Db	698	SerValProLeuAspGluGlyPheArgLysTrpThrAlaPheThrIleProSerIleAsn	717
Qy	1081	AACGAGACCCCGCATCCGCTACAGTACAACTGTCTGCCCGAGGGCTGGAAGGGCAGC	1140
Db	718	AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer	737
Qy	1141	CCGAGCATCTCCAGACGAGCATCAACAGATCTCGAGCCCTTCCGCGCGCGCAACCCC	1200
Db	738	ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaAsnPro	757
Qy	1201	GAGATCGTGATCTACCATGATCATGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1260
Db	758	GluLeuValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGly	777
Qy	1261	CAGCACCGCCCAAGATCGAGGAGCTCGCAAGACCTGTGTGTGTGTGTGTGTGTGTGT	1320
Db	778	GlnHisArgAlaLysIleGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThr	797
Qy	1321	CCGACAGAAGACACCAAG	1380
Db	798	ProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisPro	817
Qy	1381	GACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGAGAGAGAGAGAGAG	1440
Db	818	AspLysTrpThrValGlnProIleGlnLeuProGlnLysGluSerTrpThrValAsnAsp	837
Qy	1441	ATCCAGAAGCTGTGTGGCAAGCTGAACTGGGCCCAAGCAGATCTACCCCGCATCAAGTG	1500
Db	838	IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal	857
Qy	1501	CGCCAGCTGTGCAGCTGT	1560
Db	858	ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThr	877
Qy	1561	GAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGC	1620
Db	878	GluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluProValHisGly	897
Qy	1621	GTGTACTACGACCCCGACAGGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1680
Db	898	ValTyrTrpAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGln	917
Qy	1681	TGGACCTACAGATCTACAGAGAGCCCTTCAGAACCTGAAGACCGGCGAGTGCAGCAG	1740
Db	918	TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys	937
Qy	1741	ATGCGCACCCCGCACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAGAAAGTGC	1800
Db	938	MetArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla	957
Qy	1801	ATGAGAGAGCTGT	1860
Db	958	MetGluGlyIleValIleTrpGlyLysValProLysPheArgLeuProIleGlnLysGlu	977
Qy	1861	ACCTGGGAGACCTGGTGGACCGACTACTTGGCAGCGCCACCTGGATCCCCGAGTGGAGTTC	1920





Db 418 LysAspCysThrGluArgGlnAlaAsnPheArgGluIleLeuAlaPheProGlnGly 437  
Qy 253 AAGCCCGCGAGTTCCTCCAGCAGACAGAACCGCGCCCAACAGCCCAACAGCGCGAGCTG 312  
Db 438 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrAsnGlyGluLeu 457  
Qy 313 CAGGTGGCGCGGAGCAACACCCCGCAGCAGCGCCCGCGCGCGAGCGCGAGCGCCCTGAAC 372  
Db 458 GlnValArgArgAsnAsnProArgSerLysThrGlyValGluArgGlnGlyThrLeuAsn 477  
Qy 373 TTCCCCAGATCACCTGTGTGCAGCGCCCTGTGTGAGCATCAAGTGGGGCGGCAGATC 432  
Db 478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle 497  
Qy 433 AAGGAGCCCTGTGTGACACCGCGCGCGACACACCGTGTGTGAGAGAGCTGCCCC 492  
Db 498 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517  
Qy 493 GGCAGGTGGAGCAACAGATGATCGCGGGGATCGCGGGCTTCATCAAGGTGGCGCATAC 552  
Db 518 GlyLysTrpLysProLysMetIleGlyIleGlyGlyPheIleLysValArgGlnTyr 537  
Qy 553 GACCAGATCTGTGAGATCTCGGCAAGAGCCATCGCACCGCTGATCGGCCCC 612  
Db 538 AspGlnIleLeuIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyPro 557  
Qy 613 ACCCCCTGAAATCATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAATTC 672  
Db 558 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 577  
Qy 673 CCATATGAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGGCCCAAG 732  
Db 578 ProfileSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597  
Qy 733 GTGAGCAGTGGCCCTGACCGGAGGAGNATCAAGCCCTGACCGCGCATCTGCGAGAG 792  
Db 598 ValLysGlnTrpProLeuThrGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617  
Qy 793 ATGAGAGAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACAAACACCCCGTG 852  
Db 618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 637  
Qy 853 TTGCGCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGTGGACTTCGCGAGCTG 912  
Db 638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657  
Qy 913 AACAGCGCACCCAGGACTTCTGGAGGTGCTGGAGTGGCGATCCCGCACCCCGCGGCTG 972  
Db 658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677  
Qy 973 AAGAAGAAGAGAGCGTGCCTGACGCTGGCGGACGCGCTACTTTCAGCGTCCCGCTG 1032  
Db 678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 697  
Qy 1033 GACGAGGACTTCGCAAGTACACCGCTTCACCATCCCGCAGCATCAACACGAGACCCCC 1092  
Db 698 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717  
Qy 1093 GGCATCCGCTACCATGATACAGCTGCTCCCGAGGGCTGGAGGGCAGCCCCCAGCATCTTC 1152  
Db 718 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 737  
Qy 1153 CAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 738 GlnSerSerMetThrArgIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 757  
Qy 1213 TACCAGTACATGACGACCTGTACGTGGCGGACGACCTGGAGATCGCGCCAGCACCGCGCC 1272  
Db 758 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 777  
Qy 1273 AAGATCAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCACCAACCCCGCAAGAG 1332

Db 778 LysIleGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797  
Qy 1333 CACCAGAAAGAGCCCCCTTCTGTGTGGATGGCTACGAGCTGCACCCCGCAAGTGGACC 1392  
Db 798 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 817  
Qy 1393 GTGCAGCCCATCGAGCTGCCGAGAGGAGGAGCTGGACCGTGAACGACATCCAGAGCTG 1452  
Db 818 ValGlnProIleGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 837  
Qy 1453 GTGGCGAAGCTGAATCGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTC 1512  
Db 838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 857  
Qy 1513 AAGTGTGTGCGCGCGCCAAAGCCCTGACCGACATCGTGGCCCTGACCGAGAGCCGAG 1572  
Db 858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 877  
Qy 1573 CTGGAGCTGCGCGAGAACCCGCGAGATCTCTGCGCGAGCCCTGCGACGGGCTGTACTACG 1632  
Db 878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 897  
Qy 1633 CACAGCAAGGACCTGTGTGGCGGAGATCCAGAAAGCAGGSCCAGCACAGTGCATTAACCAG 1692  
Db 898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnLysAspAspGlnTrpThrTyrGln 917  
Qy 1693 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCCGCGCAAGTACGCGCAAGATCGCACCGCC 1752  
Db 918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937  
Qy 1753 CACACCAACAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAAGATCGGCATGGAGAGCATC 1812  
Db 938 HisThrAsnAspValLysGlnLeuValGlyAlaValGlnLysIleAlaMetGluGlyIle 957  
Qy 1813 GTGATCTGGCGGAGAGCCCGCAAGTCTCGCTGCCCATCCAGAAAGGAGACCTGGGAGACC 1872  
Db 958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 977  
Qy 1873 TGTGGACCGACTACTGCGAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1932  
Db 978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPheValAsnThrPro 997  
Qy 1933 CCGCTGTGAAGCTGTGTGTACAGCTGAGAGAGGAGCCCATCATCTGCGCGCGAGACCTTC 1992  
Db 998 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyValGluThrPhe 1017  
Qy 1993 TAGTGTGACGCGCGCGCAACCGGAGACCAAGATCGCAAGCGCGCTACGTGCACCGAC 2052  
Db 1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037  
Qy 2053 CGGGCGCGCAGAGATCTGTGAGCTGACCGAGACCAACCAACAGAAAGACCGAGCTGCAG 2112  
Db 1038 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057  
Qy 2113 GCATTCAGCTGCGCCCTGCGAGGACGAGCGGCGAGGTGAACATCTGTGACCGAGCCGAG 2172  
Db 1058 AlaIleGlnIleAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1077  
Qy 2173 TAGCCCTGGGATCATCTCAGGCGCCAGCCCGCAAGAGCGAGCGAGCTGGTCAACCCAG 2232  
Db 1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097  
Qy 2233 ATCATTCAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCCCGCCCAACAG 2292  
Db 1098 IleIleGluGlnLeuIleLysGluArgIleTyrLeuThrTrpValProAlaHisLys 1117  
Qy 2293 GGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAAGGGCATCCCGCAAGGTGCTG 2352  
Db 1118 GlyIleGlyAsnGluGlnValAspLysLeuValSerAsnGlyIleArgLysValLeu 1137  
Qy 2353 TTCTCGACCGCATCGAT 2370  
Db 1138 PheLeuAspGlyIleAsp 1143



Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAGCGCACCAGGACTTCTGGAGGTGAGCTGGGCATCCGCCACCCCGCGGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAAGAAGCGTGACCGTCTGTGACGTGGCGACGGCTACTTTCAGCGTCCCGT 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGGACTTCCCAAGTACACCGCTTCCACCATCCCGACATCAACAACGAGACCC 1092  
Db 700 AspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719  
Qy 1093 GGCACTCCGCTACCAAGTACAAGCTGCTGCCAGGGCTGGAGGGCAGCCCGACGATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy 1153 CAGAGCATGATCAAGATCCTGAGCCCTTCCGGCCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759  
Qy 1213 TACCAGTACATGGACGACTCTAGCTGGGACGACCTGGAGATCGGCGCACCGCGCC 1272  
Db 760 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1273 AAGATCGAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCCACCCCGCAAGAAG 1332  
Db 780 LysIleGluLeuLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799  
Qy 1333 CACAGAAGAGCCCTTCTCTGTGATGGCTACGAGCTGCACCCCGCAAGTGGACC 1392  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1393 GTGCAGCCATCGAGCTGCCGACAGAGAGCTGGACCGTGAACGATCCCAAGCTG 1452  
Db 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1453 GTGGCAAGCTGAACTGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTC 1512  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
Qy 1513 AAGTGTGCGCGCGCCAGGCTCTGACGACATCGTGCCCTGACGAGAGGCGCAG 1572  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879  
Qy 1573 CTGAGCTGGCGAGAACCGGAGATCTCGCGCAGCCCGTGACGGCGTGTACTAGAC 1632  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLysLysGluProValHisGlyValTyrTyrAsp 899  
Qy 1633 CCCAGCAAGGACCTGTGGCGGAGATCCAGAGCGGCGCACACGAGTGGACCTACGAG 1692  
Db 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1693 ATCTACGAGGACCTTTCAGAACTGAGACCGCGCAAGTACGCGCAAGATCGCGACCGCC 1752  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1753 CACACCAACGCTGAGACGCTGACCGAGCGCTGCAGAGATCGCCATCGAGAGCATC 1812  
Db 940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959  
Qy 1813 GTGATCTGGGCGAAGACCCCAAGTCTCGCTCGCCATCCAGAGGAGAGCTGGGAGACC 1872  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1873 TGTGGACCGCACTACTGCGAGGCCACTGATCCCGAGTGGGAGTTCGTGAACACCC 1932  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1933 CCCTGTGAGCTGTGGTACGACTGAGAGAGCCCATCATCGCGCGCGAGACCTTC 1992  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy 1993 TACGTGACCGCGCCGCCAACCGGACACCGGACATCGCAAGCGCGGTACGTGACCCAG 2052  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2053 CGGGCGCGGACAGATCGTGGAGCTGACCGGACACCGACACCGACGAGCTGCGAG 2112  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
Qy 2113 GCCATCCAGCTGGCCCTGCGAGGACGCGGACGAGGTGAACATCTGTCACCGACGAG 2172  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079  
Qy 2173 TACGCTCGGCTGATCATCGAGCCCGCCGACCAAGCGCGAGCGAGCTGCTGAACCCAG 2232  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
Qy 2233 ATCATCCAGCAGCTGATCAAGAGGAGAAGGTGTACCTGAGCTGGTGGCGCCGACCAAG 2292  
Db 1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy 2293 GGCATCCGCGCGCAACGAGCAGATCGCAAGCTGTGAGAGGCGCATCCGCAAGTGTCTG 2352  
Db 1120 GlyIleGlyGlnAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy 2353 TTCCTGACCGCATCGAT 2370  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 8  
Q6X4R4 ID Q6X4R4 PRELIMINARY; PRT; 1430 AA.  
AC Q6X4R4; AT Q6X4R4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Gag-pol fusion polyprotein (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/089922203771881220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133 (2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Harris M.E., Birx D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AY255823; AAF76521.1; -;  
DR HSP; P12497; I89D.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.

Qy	313	CAGGTGGCGGCACAAACCCCGCACGAGGCGCGCGCGAGCGC-----CAG	360
Db	459	GlnValArgGlyAspAsnProLeuSerGluAlaGlyThrGluArgProGlyThrLeuGln	478
Qy	361	GGCACCTGAACCTTCCCCAGATACACCTGTGGCAGCGCCCTCGTGAGCATCAAGGT	420
Db	479	GlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleValIle	498
Qy	421	GGCGGCCAGATCAAGAGGCCCTGTGGACACCGCGCCGACACACCGTGTGGAGGAG	480
Db	499	GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu	518
Qy	481	ATGAGCCTGCCCGCAAGTGAAGCCCAAGATCATCGCGGCATCGCGGCTTCATCAAG	540
Db	519	IleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys	538
Qy	541	GTCCGCCAGTACCACACAGATCCTGATCGAGATCTGCGGCAAGGACATCGGCACCGT	600
Db	539	ValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlySerVal	558
Qy	601	CTGATCGGCCCAACCCCGTGAAACATCATCGGCCGCAACATGTGACCCAGCTGGGCTGC	660
Db	559	LeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCys	578
Qy	661	ACCTGAACHTTCCCATCAGCCCCATCGAGACCGTCCCGTGAGCTGAAGCCGCGCATG	720
Db	579	ThrLeuAsnPheProIleSerProIleGluThrIleProValLysLeuLysProGlyMet	598
Qy	721	GACGGCCCCAAGGTGAAGCAGTGGCCCTGACCGAGGAGAAATCAAGGCCCTGACCGCC	780
Db	599	AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAla	618
Qy	781	ATTGCGAGGAGATGAGAAAGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTAC	840
Db	619	IleCysGluAspMetGluArgGluGlyLysIleSerLysIleGlyProGluAsnProTyr	638
Qy	841	AACACCCCGTGTGCCCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGAC	900
Db	639	AsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAsp	658
Qy	901	TTCCGGGAGTGAACAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCAC	960
Db	659	PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis	678
Qy	961	CCCGCGGCTGAAGAAGAAAGAGCGTGACCGTCTGTGACGTGGCGAGCGCTACTTC	1020
Db	679	ProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe	698
Qy	1021	AGCGTCCCTGACGAGGACTTCCGCAAGTACACCGCCTTACCACCTCCCGAGCATCAAC	1080
Db	699	SerValProLeuAspGluAsnPheArgLysTyrThrAlaPheThrIleProSerThrAsn	718
Qy	1081	AACGAGACCCCGGCATCCGCTACCAAGTACAACGTGTCTGCCAGGCTGGAAGGCGAGC	1140
Db	719	AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGlySer	738
Qy	1141	CCGAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCC	1200
Db	739	ProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro	758
Qy	1201	GAGATCGTGATACCAAGTACATGGACGACTGTACTGGGCGACGACCTGGAGATCGCG	1260
Db	759	GluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly	778
Qy	1261	CAGCACCGCCCAAGATCGAGGAGCTCGCAAGCACTGTCTGGCTGGGCTTCACCAAC	1320
Db	779	GlnHisArgAlaLysValGluGluLeuArgGlyHisLeuLeuArgTrpGlyPheThrThr	798
Qy	1321	CCGAGCAAGACACAGAGGAGCCCCCTTCTGTGAGTGGCTACGAGCTGCACCCC	1380
Db	799	ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro	818

Qy 1381 GACAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAGAGAGCTGCAGCGTGAACGAC 1440  
 Db AsplysTrpThrValGlnThrIleGlnLeuProGluLysAspSerTrpThrValAsnAsp 838  
 Qy 1441 ATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCCAGCCAGCATCTACCCCGGATCAAGGTG 1500  
 Db IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal 858  
 Qy 1501 CGCAGCTGTGCAAGCTGCTGGCGGCGCCCAAGGCCCTGACCGCATCGTGCCTGACCC 1560  
 Db ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValAlaLeuThr 878  
 Qy 1561 GAGGAGCCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCGCGGAGCCCGTGCAGCGC 1620  
 Db GluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValIleGly 898  
 Qy 1621 GTGTACTACGACCCAGCAGGACTGTGTGGCCGAGATCCAGAGAGCGGGCCAGCAGCAG 1680  
 Db ValTyrTyrAspProSerLysAspLeuIleAlaGluLeuGlnLysGlnGlyHisAspGln 918  
 Qy 1681 TGGACCTACAGATCTTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740  
 Db TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 938  
 Qy 1741 ATGGCGACCGCCACACCAACGAGCTGAGAGCAGCTGACGAGCGCGTGCAGAGATGCC 1800  
 Db ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 958  
 Qy 1801 ATGAGAGACCTCGATCTGGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1860  
 Db MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu 978  
 Qy 1861 ACCTGGGAGACCTGTGTGACCGCACTACTTGGCAGCGCCACCTGGATCCCGAGTGGAGTTC 1920  
 Db ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 998  
 Qy 1921 GTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGAGCCCATCATCGGC 1980  
 Db ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 1018  
 Qy 1981 GCCGAGACCTTCTACGTGGCGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGC 2040  
 Db AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1038  
 Qy 2041 TACGTGACCGACCGCGCGCGCAGAGATCTGTGACCTGACCGAGACCCACCAACGAGAAG 2100  
 Db TyrValThrAspLysGlyArgGlnLysValValSerLeuAsnGluThrThrAsnGlnLys 1058  
 Qy 2101 ACCGAGTGCAGGCGCATCTGAGCTGGCGCTGCAGCAGCGCGCAGCGGAGGAGCATCGTG 2160  
 Db ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1078  
 Qy 2161 ACCGACAGCAGTACCGCGCTGGCGCATCATCCAGCGCCGAGCGCGCAGAGCGAGCGAG 2220  
 Db ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerAspSerGlu 1098  
 Qy 2221 CTGTGTGAACAGATCATCTGACGACGCTGATCAAGAGGAGAGGTGTACCTGAGCTGGGTG 2280  
 Db IleValAsnGlnIleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpVal 1118  
 Qy 2281 CCCGCCACAGGCGATCGCGCGCAACGAGCAGATCGACAGCTGGTGGAGCAAGCGGATC 2340  
 Db ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerLysGlyIle 1138  
 Qy 2341 CGCAGCTGTCTTCTGGAGCGGATCGAT 2370  
 Db ArgLysValLeuPheLeuAspGlyIleAsp 1148

RESULT 9  
 Q9WF71  
 ID Q9WF71 PRELIMINARY; PRT: 1427 AA.  
 AC Q9WF71;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana." RT  
 RL J. Virol. 73:4427-4432(1999).  
 CC -1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110964; AAD17047.1; -;  
 DR HSP; Q76622; IHVN.  
 DR GO: GO:0019012; C:virion; IEA.  
 DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR GO: GO:0008907; F:integrase activity; IEA.  
 DR GO: GO:0008233; F:peptidase activity; IEA.  
 DR GO: GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO: GO:0003723; F:RNA binding; IEA.  
 DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO: GO:0005198; F:structural molecule activity; IEA.  
 DR GO: GO:0016740; F:transferase activity; IEA.  
 DR GO: GO:0008270; F:zinc ion binding; IEA.  
 DR GO: GO:0015074; P:DNA integration; IEA.  
 DR GO: GO:0006310; P:DNA recombination; IEA.  
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO: GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_P24.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn\_N.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR009007; Pept\_Aspatic.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR010999; Retrovirk\_matrix.  
 DR InterPro; IPR000071; Retrovirk\_p17.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR004477; RVTse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00098; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIV1MATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS00158; Zf\_CCHC; 2.  
 DR AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 1427 AA; 161326 MW; B9FC43C29B26CFA2 CRC64;

Alignment Scores:

Pred. No.:	3,96e-151	Length:	1427	QY	973	AAGAGAGAGAGAGCGCTGACCGCTGTGGACGCGCGGCGAGCGCCCTTACTTACGCGTGCCTG	1032
Score:	3973.50	Matches:	731	Db	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
Percent Similarity:	97.46%	Conservative:	35	QY	1033	GACGAGGACTTCGCGAAGTACACCGCTTACCATCCAGCATCAACACGAGACCC	1092
Best Local Similarity:	93.00%	Mismatches:	19	Db	700	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro	719
Query Match:	86.70%	Indels:	1	QY	1093	GGCATCCGCTACCACTACACGCTGCTGCCCGGCTGGAAGCGGAGCCCGACCATCTTC	1152
DB:	2	Gaps:	1	Db	720	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrPheSerProAlaIlePhe	739
US-09-610-313B-30 (1-2469) x Q9WF71 (1-1427)				QY	1153	CAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATC	1212
				Db	740	GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
				QY	1213	TACCATCATGAGCAGCATCTGACGTGGGAGGAGCTGGAGATCGGCGAGACCGCGCC	1272
				Db	760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
				QY	1273	AAGATCGAGGAGCTGCGCAAGCACCTGCTGGCTGGGCTTACACACCCCGCAACAAG	1332
				Db	780	LysIleGluGluLeuArgGlnHisLeuLeuLysTyrGlyLeuThrThrProAspLysLys	799
				QY	1333	CACCAGAGGAGCCCTTCTCTGTGGATGGGCTACGAGCTGCACCCCGCAAGTGGACC	1392
				Db	800	HisGlnLysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThr	819
				QY	1393	GTGCACCCATCGAGCTGCGCGAGAGGAGCTGACCGTGAACGATCCAGACATCCAGAGCTG	1452
				Db	820	ValGlnProIleGlnLeuProAspLysAspSerTyrThrValAsnAspIleGlnLysLeu	839
				QY	1453	GTGGGCAAGCTGAACCTGGCGCCAGATCTACCCCGCATCAAGTGGCCAGCTGTGC	1512
				Db	840	IleGlyLysLeuAsnThrPheSerGlnIleTyrProGlyIleArgValLysHisLeuCys	859
				QY	1513	AAGCTCTGCGCGCGCCCAAGCCCTGACCGACATCGTCCCTGACCGAGAGGCGGAG	1572
				Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu	879
				QY	1573	CTGGAGCTGCGGAGAACCGCGAGATCTTGGCGGAGCCCGTGCACCGCTGTACTAGC	1632
				Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
				QY	1633	CCAGCAAGACCTGTGTGGCGGAGATCCAGAACGAGCGCCACACACAGTGGACTAC	1692
				Db	900	ProSerLysAspLeuIleAlaGluAlaGlnLysGlnGlyHisAspGlnThrTyrGln	919
				QY	1693	ATCTACCGAGAGCCCTTCAAGAACCTGAAGACCGCAAGTACCGCAAGATCGGACCGCC	1752
				Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
				QY	1753	CACACCAACGATGAGCGAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAGAGATC	1812
				Db	940	HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle	959
				QY	1813	GTGATCTGGGGCAAGACCCCAAGTTCCGCTCCCATCCAGAACGAGACCTGGGAGACC	1872
				Db	960	ValIleTyrLysThrProLysPheArgLeuProIleGlnLysAspThrTyrGluThr	979
				QY	1873	TGGTGGACCGACTACTGCGAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCC	1932
				Db	980	TyrThrAspTyrTyrGlnAlaThrTyrIleProGluTyrGluPheValAsnThrPro	999
				QY	1933	CCCTGTGTGAGCTGTGTACAGCTGGAGAGGCCCATCATCGCGCGCGAGACCTTC	1992
				Db	1000	ProLeuValLysLeuTyrTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe	1019
				QY	1993	TAGTGTGACCGCGCCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACTGTACCGAC	2052
				Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp	1039
				QY	2053	CGGGCGCGGAGAGATCGTGAGCCTGACCGAGACCAACCAACAGAGACCGAGCTGAG	2112

```

Db 1040 LysGlyArgGlnLysIleValSerLeuAenGluThrThrAenGlnLysAlaGluLeuGln 1059
Qy 2113 CCCATCCAGCTGGCCCTCAGCAGCAGCGCAGCGAGGTGAACATCGTGACCGCAGCCAG 2172
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAenIleValThrAspSerGln 1079
Qy 2173 TACGCTCGGCATCATCCAGGCGCAGCCGACAGAGCGAGCGAGCTGGTGAACCAAG 2232
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099
Qy 2233 ATCATCCAGCAGCTGATCAAGAAGGAGAGCTGTACCTGAGCTGGTGGCCGCCCAAG 2292
Db 1100 IleIleGluGlnLeuIleLysGlyGluArgValTyrLeuSerTrpValProAlaHisLys 1119
Qy 2293 GGCAATCGCGCCGCAACGAGCAGATCCACAAGCTGGTGAAGGCGATCCGCAAGGTGCTG 2352
Db 1120 GlyIleGlyGlnAenGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
Qy 2353 TTCCTGAGCGCATCGAT 2370
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 10
Q9WFAO PRELIMINARY; PRT; 1427 AA.
AC Q9WFAO;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OC NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana."
RL J. Virol. 73:4427-4432(1999).
CC -!- SIMILARITY: Belongs to peptidase family A2.
DR EMBL: AF110978; AAD17162.1; -.
DR PIR: S49086; S49086.
DR HSP; Q70622; IHVN.
DR GO: GO:0019012; C:virion; IEA.
DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0008907; F:integrase activity; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004523; F:ribonuclease H activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0015074; F:DNA integration; IEA.
DR GO: GO:0006310; F:DNA recombination; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO: GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO: GO:0016032; P:viral life cycle; IEA.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR001037; Integrase C.
DR InterPro: IPR003308; Integrase Zn N.
DR InterPro: IPR001995; Peptidase A2.
DR InterPro: IPR009007; Pept Aspartic.
DR InterPro: IPR001969; Pept_Asp_AS.
DR InterPro: IPR010999; Retrovir_matrix.
DR InterPro: IPR000071; Retrovir_p17.
DR InterPro: IPR008916; Retrov_capsid_C.

```

```

DR InterPro: IPR008919; Retrov_capsid_N.
DR InterPro: IPR002156; RNaseH.
DR InterPro: IPR001584; Rve.
DR InterPro: IPR000477; RVTse.
DR InterPro: IPR010659; RVT_connect.
DR InterPro: IPR010661; RVT_thumb.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00540; Gag_p17; 1.
DR Pfam: PF02022; Integrase_Zn; 1.
DR Pfam: PF00075; RNaseH; 1.
DR Pfam: PF00665; Rve; 1.
DR Pfam: PF00077; RVP; 1.
DR Pfam: PF00078; RVT_1; 1.
DR Pfam: PF06815; RVT_connect; 1.
DR Pfam: PF06817; RVT_thumb; 1.
DR Pfam: PF00098; zf_CCHC; 2.
DR PRINTS: PR00939; C2HCZNFINGER.
DR PRINTS: PR00234; HIVMATRIX.
DR SMART: SM00343; Znf_C2HC; 2.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS0175; ASP_PROT_RETROV; 1.
DR PROSITE: PS0158; ZF_CCHC; 2.
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 160503 MW; 24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.: 6e-151 Length: 1427
Score: 3969.00 Matches: 731
Percent Similarity: 97.46% Conservative: 35
Best Local Similarity: 93.00% Mismatches: 20
Query Match: 86.60% Indels: 0
DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x Q9WFAO (1-1427)
Qy 13 ATGCCCGAGCGCATGAGCGCCAGCGCCAGCCAGCCCAACATCTCTGTATCGACGCGAGCAACTTC 72
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnThrAsnIleMetIleGlnArgSerAsnPhe 379
Qy 73 AAGGCGCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGGCCACATCGCCCGC 132
Db 380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399
Qy 133 AACTGCGCGCGCCCGCCAGAGAGGGCTGCTGGAGTGGCGGAGGAGGGCCAGCATG 192
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisIleMet 419
Qy 193 AAGGACTGCGCCAGCGCCAGCGCCAGCCAACTTCTTCGCGAGGACCTCTTCCCCCAGGCG 252
Db 420 LysAspCysThrGluArgGlnAlaAspPhePheArgGluAenLeuAlaPheProGlnGly 439
Qy 253 AAGGCGCGGAGTTCCTCCAGCGAGCAGAAACCGCGCCCAACAGCCCGCCGAGCGTG 312
Db 440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459
Qy 313 CAGGTGCGCGCGCAGACACCCCGCAGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479
Qy 373 TTCCTCCAGATCACCTCTGTGCGAGCGCGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 432
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnVal 499
Qy 433 AAGGAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 492
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 519
Qy 493 GGCNAGTGGAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 552
Db 520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539

```







1273 AGATCGAGGAGTGGCAAGCACCTGCTCGCTGGGGCTTACACCCCGGACGAAG 1332  
 Db |||||LysileGluLeuLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProTyrLysLys 799  
 1333 CACCAGAGGAGCCCTCTCTGTGGATGGGTACGAGCTGACCCCGCAAGTGGACC 1392  
 Db |||||HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
 1393 GTGAGCCCATCGAGTGGCCGAGAGGAGAGCTGGACCTGAACGACATCCGAGCTG 1452  
 Db |||||ValGlnProIleGlnLeuProAspLysAspSerTrpThrValHisAspIleGlnLysLeu 839  
 1453 GTGGGCAAGCTGAACCTGGGCGCCAGATCTACCCGGCATCAAGTGGCCAGCTGTGC 1512  
 Db |||||ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysHisLeuLys 859  
 1513 AACTGTGTGGCGCGCCCAAGGCCCTGACGACATCGTGCCTCGACCGAGAGGCGGAG 1572  
 Db |||||LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu 879  
 1573 CTGGAGTGGCCGAGAACCCGAGATCTCTGGGAGCCCTGACCGGCTGACTAGCAC 1632  
 Db |||||LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 899  
 1633 CCAGCAAGGACCTGTGGCGCGAGATCCACAAGCAGGCGCCACGACGATGGACTAC 1692  
 Db |||||ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrGln 919  
 1693 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCC 1752  
 Db |||||IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
 1753 CACACCAAGCTGAAGAGCTGACCGAGCGCTGCAGAAGATCGCCATCGGAGGATC 1812  
 Db |||||HisThrAsnAspValLysGlnLeuThrGluValGlnLysIleThrMetGluSerIle 959  
 1813 GTGATCTGGGCGAAGCCCAAGTCTCGCTGCCATCCAGAGGAGACTCGGAGACC 1872  
 Db |||||ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979  
 1873 TGGTGAACCGACTACTGGCAGGCGACCTGGATCCCGAGTGGGAGTTCGTGAACACCCC 1932  
 Db |||||TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
 1933 CCCTGTGTGAAGCTGTGGTACAGCTGGAGAAGAGCCCATCATCGCGCGCGAGACCTTC 1992  
 Db |||||ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
 1993 TAGTGGACCGCGCCGCAACCGGAGACCAAGATCGCAAGCGCGCTACGTGACCGAC 2052  
 Db |||||TyrValAspGlyAlaAlaAsnArgGluThrLysLysAlaGlyTyrValThrAsp 1039  
 2053 CGGCGCGCGCAGAGATCGTGAGCTGTACCGGACCAACAGAGAGACCGAGCTGCAG 2112  
 Db |||||LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln 1059  
 2113 GCATTCAGCTGGCCCTGTGAGGACGCGGAGCGAGGTGAACATCTGTGACCGACGAC 2172  
 Db |||||AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
 2173 TAGCCCTGTGGGCATCATCTCAGGCGCCGCGCAGCAAGAGCGAGCGAGCTGGTGAAC 2232  
 Db |||||TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
 2233 ATCATCGAGGAGCTGTGATCAAGAGGAGAGTGTACTGAGCTGGTGGCGCCGCCAAG 2292  
 Db |||||IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
 2293 GGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGAGCAGGCGCATCCGCGAGGTCTG 2352  
 Db |||||GlyIleGlyLysAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

2353 TTCTGCGACGCGATCGAT 2370  
 Db |||||PheLeuAspGlyIleAsp 1145  
 RESULT 12  
 Q9WF77 PRELIMINARY; PRT; 1437 AA.  
 AC Q9WF77  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana.";  
 RL J. Virol. 73:4427-4432(1999).  
 CC -1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AF110967; AAD17072.1; --  
 DR HSSP; P24736; INCP.  
 DR GO; GO:0019012; C:virion; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:integrase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003364; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0015074; P:DNA integration; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn\_N.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR009007; Peptidase.  
 DR InterPro; IPR001969; Peptidase.  
 DR InterPro; IPR010999; Retroviral\_matrix.  
 DR InterPro; IPR000071; Retroviral\_p17.  
 DR InterPro; IPR008916; Retroviral\_capsid\_C.  
 DR InterPro; IPR008919; Retroviral\_capsid\_N.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00098; zfc-CHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIVMATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.





480	Db	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLeuValSerIleLeuValGlyGlnIle	499
433	Qy	AAGGAGGCGCTGCTGGACAACCGGGCGCCAGCACACCGTGTGGAGGAGATGAGCTGCC	492
500	Db	ArgGluAlaLeuLeuAspThrGlyAlaAspThrValLeuGluAspIleAsnLeuPro	519
493	Qy	GGCAAGTGGAAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCAAGTGGCGCAGTAC	552
520	Db	GlyLysTrpLysProLysMetIleGlyGlyPheIleLysValArgGlnTrp	539
553	Qy	GACCAGATCCTGATCGAGATCTCGGGCAAGAAGCCATCGGCACCGCTGTGATCGGCCCC	612
540	Db	GluGlnIleLeuIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyPro	559
613	Qy	ACCCCGTGAACATCATCGGCGCGAAACATGTGTGACCCAGCTGGGCTGCACCTCAACTTC	672
560	Db	ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe	579
673	Qy	CCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATGAGAGCGGCCCAAG	733
580	Db	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
733	Qy	GTGAAGCAGTGGCCCTCACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTCGAGGAG	792
600	Db	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp	619
793	Qy	ATGAGAGAAGGGCAAGATCACCAAGATCGGCCCGAGAACCCCTACAACACCCCGCTG	852
620	Db	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal	639
853	Qy	TTCGCCATCAAGAAAGAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGAGCTG	912
640	Db	PheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
913	Qy	AACAAGCGCACCGAGACTTCTGGAGGTGCGAGCTGGGCATCCCGCACCGCGCGCTG	972
660	Db	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679
973	Qy	AAGAAGAAGAAGACGTGACCGTCTGTGACGTGGCGCAGCGCTACTTTCAGCGTCCCGCTG	1032
680	Db	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
1033	Qy	GACGAGACTTCCGCAAGTACACCGCTTCACCATCCCCAGCATCAACAAGAGAGACCCC	1092
700	Db	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro	719
1093	Qy	GGCATCGCTACAGTACAGTCTGCTGCCCGAGCGCTGGAGGCGACGCCAGCATCTTC	1152
720	Db	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	739
1153	Qy	CAGACGACATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCGAGATCGTGATC	1212
740	Db	GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
1213	Qy	TACCAGTACATGGACGACCTGTACGTGGGCGAGACCTGGAGATCGGCCAGCACCGGCC	1272
760	Db	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
1273	Qy	AAGATCAGGAGCTGCCAAGCACCCTGCTGGCTGGGCTTCAACACCCCGCACAGAAG	1332
780	Db	LysIleGluLeuLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys	799
1333	Qy	CACCAGAAGGAGCCCCCTTCTGTGATGGGCTACGAGCTGCACCCCGCACAGTGGACC	1392
800	Db	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
1393	Qy	GTGCAGGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGGTGAACGACATCCCAAGCTG	1452
820	Db	ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu	839
1453	Qy	GTGGGCAAGCTGAAGCTGGCGCAGCCAGATCTACCCCGGATCAAGTGGCGCAGCTGTC	1512
840	Db	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValArgHisLeuLysCys	859

QY	1513	AA	CTGTGTGGCGCGCGCAAGGGCCCTGACCGACATCGTGGCCCTGACCGAGGAGCGCCGAG	1572
DB	860	Lys	LeuLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu	879
QY	1573	CT	GAGAGTGGCGGAGAAACCGCGAGATCTCTGGCGAGCCGCTGCACGCGGTGTACTACGAC	1632
DB	880	Leu	gluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
QY	1633	CC	CAGCAAGACCTGTGTGGCCGAGATCCAGAAGCGAGGCCACCAACCACTGCAGCTGCACCTAC	1692
DB	900	Pro	SerLysAspLeuIleAlaGluValGlnLysGlnHisAspGlnTrpThrTyrGln	919
QY	1693	AT	CTTACCAGGAGCCCTTCAAGAACTCTGAAGACCGGCAAGTAGTACGCCAAGATGCGCACCGCC	1752
DB	920	Ile	TyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
QY	1753	CAC	ACCAAACAGCTGAAGCAGCTGCACGAGGCCGTGCAGAAGATCGCCATCGGAGAGCATC	1812
DB	940	His	ThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle	959
QY	1813	GT	GATCTGGGGCAAGACCCCCAAGTTCGCGCTGCCATCCAGAAGGAGACCTGGGAGACC	1872
DB	960	Val	IleTrpGlyLysThrProLysPheArgLysProIleGlnLysAspThrTrpGluThr	979
QY	1873	TGT	TGGACCGACTACTGGCAGGGCCACCTGTGATCCCCAGTGGGAGTTTCGTGAACACCCCC	1932
DB	980	Trp	TrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
QY	1933	CC	CTGTGTGAAGCTGTGTGTACCACTGGAGAGAGCCCATCATCGCGCGCGAGACCTTC	1992
DB	1000	Pro	LeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe	1019
QY	1993	TAC	GTGACCGGCGCCCAACCGCGAGACCAAGTCCGCAAGGCGCGCTACGTGACCGAC	2052
DB	1020	Tyr	ValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	1039
QY	2053	CG	GGCGCGGAGAGATCTGTGAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAG	2112
DB	1040	Lys	GlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnLysAlaGluLeuGln	1059
QY	2113	GC	ATTCAGCTGGCCTGCAAGCAGCGGAGCGAGGTGAACATCTGTGACCGCAGCAGCCAG	2172
DB	1060	Ala	IleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln	1079
QY	2173	TAC	GCCTGGGCATCATCAGGCGCCCGACCGAGACCAAGAGCGAGACGAGCTGGTGAAC	2232
DB	1080	Tyr	AlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln	1099
QY	2233	AT	CATCAGCAGCTGTATCAAGAGGAGAGGTGTACTCTGAGCTGGGTGCGCGGCCCAAG	2292
DB	1100	Ile	IleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys	1119
QY	2293	GG	CATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCCGCAAGGTGCTG	2352
DB	1120	Gly	IleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu	1139
QY	2353	TT	CCTGGACGGCATCGAT	2370
DB	1140	Phe	LeuAspGlyIleAsp	1145
RESULT 14				
Q8ADX5	PRELIMINARY; PRT; 1427 AA.			
ID	Q8ADX5			
AC	Q8ADX5			
DT	01-MAR-2003	(TrEMBLrel. 23, Created)		
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE	Gag-pol fusion polyprotein (Fragment).			
GN	Name=gag-pol;			
OS	Human immunodeficiency virus 1.			
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.			
OX	NCBI TaxID=11676;			





QY 793 ATGAGAGGAGGCAAGATCACCAAGATCGGCGCCGAGAACCCCTACACACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639  
QY 853 TTCGCCATCAAGAGAGGAGCAGCACCAAGTGGCGAGCTGGTGGACTTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
QY 913 AACAGAGCCACCCAGGACTTCTGGAGAGTGCAGCTGGGCAATCCCCACCCCGCGGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
QY 973 AAGAAGAGAGAGCGTGCCTGCTGAGCTGGGCGACGCTACTTTCAGGCTGCCCTG 1032  
Db 680 LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
QY 1033 GACGAGACTTCCCAAGTACACCGCTTACCATCCCGACATCAACCAACGAGACCCCC 1092  
Db 700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GGCAATCCGCTACCAAGTACACGCTGCTGCCAGGGCTGGAGGCGACCCCGAGCATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
QY 1153 CAGAGCAGCATGACCAAGATCCTGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759  
QY 1213 TACCAGTACATGGACGACTCTAGTGGGCGAGCAGCTGGAGATCGCGCAGCACCGCGCC 1272  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779  
QY 1273 AAGATCAGGAGCTGCGCAGCACCTGCTGCGCTGGGGCTTCCACACCCCGCAAGAAG 1332  
Db 780 LysIleGluGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
QY 1333 CACAGAGAGGAGCCCTTCTGCTGTGATGGGTACGAGCTGCACCCCGCAAGTGAC 1392  
Db 800 HisGlnLysGluProPhePheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
QY 1393 GTGACAGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCACAGCTG 1452  
Db 820 ValGlnProIleGlnLeuProGluLysAspSerTrpValAsnAspIleGlnLysLeu 839  
QY 1453 GTGGCGCAAGCTGAATCGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTC 1512  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 859  
QY 1513 AAGCTGTGCGCGCGCAAGGCGCTGACCGACATCGTGGCCCTGACCGAGAGCGCGAG 1572  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluLysAlaGlu 879  
QY 1573 CTGAGCTGGCGAGACCGGAGATCTGCGCGAGCCCGTGCAGCGGCTGTACTACGAC 1632  
Db 880 LeuGluLeuAlaGluAsnArgGluLeuLysGluProValHisGlyValTyrTyrAsp 899  
QY 1633 CCCAGCAGGACCTGTGGCGCGAGATCCAGAAGCAGGCGCCACGACCATGAGCTACACAG 1692  
Db 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpTyrGln 919  
QY 1693 ATCTACAGGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCC 1752  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939  
QY 1753 CACACCAACGCTGAGCAGCTACCGAGGCGCTGCAGAGATCGCATGGAGAGCATC 1812  
Db 940 HisthrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959  
QY 1813 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCATCCAGAAGAGAGACTGGGAGAC 1872  
Db 960 IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979  
QY 1873 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCC 1932

Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY 1933 CCCCTGTGAAGCTGTGTACCTACCGTGGAGAAGAGCCCATCATCGCGCGCAGACCTTC 1992  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019  
QY 1993 TACGTGACGGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2052  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039  
QY 2053 CGGGCGCGCAGAGAGTCTGAGCTCAGCGAGACCCAGACCAAGAGACCGAGCTGCAG 2112  
Db 1040 ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
QY 2113 GCATCCAGCTGGCCCTGCAGGAGCAGCGGAGGAGGAGTGAACATCGTGCACCGACCGCAG 2172  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValThrAspSerGln 1079  
QY 2173 TAGCCCTGGGCATCTCCAGGCGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCG 2232  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099  
QY 2233 ATCATCCAGCAGCTGATCAAGAGGAGAGTGTACCTGAGTGGTGGCGCCCGCAAGTGTG 2292  
Db 1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2293 GGATCCGCGCGCACGAGCAGATCGACAGCTGTGTGAGCAAGGCGCATCCGCAAGTGTG 2352  
Db 1120 GlyIleGlyGlnAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
QY 2353 TTCTCGACCGCATCGAT 2370  
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 15  
Q9WF79  
ID Q9WF79 PRELIMINARY; PRT; 1425 AA.  
AC Q9WF79;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110969; AAD17087.1; --.  
DR HSP; P24736; INCP.  
DR GO; GO:001912; C:viroin; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR001037; Integrase\_C.  
 DR InterPro; IPR003308; Integrase\_Zn\_N.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR009007; Pept\_Aspartic.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR010999; Retroviral\_matrix.  
 DR InterPro; IPR000071; Retroviral\_p17.  
 DR InterPro; IPR008916; Retroviral\_capsid\_C.  
 DR InterPro; IPR008919; Retroviral\_capsid\_N.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF06070; Gag\_p24; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00096; zf-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.  
 DR PRINTS; PR00234; HIV1MATRIX.  
 DR SMART; SM00343; Znf\_C2HC; 2.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS01075; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS00158; ZF\_CCHC; 2.  
 KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 1425 AA; 160569 MW; 6DC8A1AEF3C8BDA7 CRC64;

Alignment Scores:  
 Pred. No.: 6.55e-150 Length: 1425  
 Score: 3943.00 Matches: 731  
 Percent Similarity: 97.33% Conservative: 34  
 Best Local Similarity: 93.00% Mismatches: 19  
 Query Match: 86.04% Indels: 2  
 DB: 2 Gaps: 1

US-09-610-313B-30 (1-2469) x Q9WF79 (1-1425)

QY 13 ATGGCCGAGCCATGAGCCAGGCCACCAACATCTGATGCGAGCGCAGCACTTC 72  
 Db 360 LeuAlaGluAlaMetSerGlnAlaAsnSerAsnMetGlnLysAsnAsnPhe 379

QY 73 AAGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGCGCAAGGAGGGCCACATCGCCGC 132  
 Db 380 LysGlyProThrArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399

QY 133 AACTGCGCGCCCCCCAGAGAGGGCTGTGGAAGTGGCGGCAAGAGGGCCACCATG 192  
 Db 400 AsnCysLysAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419

QY 193 AAGGACTGACCGAGCCGAGGCCCACTCTTCGCGAGGACCTGGCTTCCCCCAGGGC 252  
 Db 420 LysAspCysThrGluArgAlaAsnPhePheArgGluAlaPheProGlnGly 439

QY 253 AAGCCCGCGAGTTTCCCGAGCGAGCAGAACCGCGCCCAACAGCCCCCAGCCGAGCTG 312  
 Db 440 GluAlaArgGluPheProSerGluGlnThrGlyAlaAsnSerProThrSerArgLysLeu 459

QY 313 CAGTGGCGGGCGACAAACCCCGGAGCGCGCGCGCCGAGCGCCAGGGCACCTGAAC 372  
 Db 460 -----ArgGlyAsnAsnProCysSerGluAlaGlyAspGluArgGlnGlyThrLeuAsn 477

QY 373 TTCCCCCGAGATCACCTGTGGCAGCCGCCCTGTGTGAGCATCAAGGTGGCGGCCAGATC 432  
 Db 478 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle 497

QY 433 AAGGAGCCCTGTGTGACACCCGGCGCGAGGACACCGTGTGTGAGAGATGAGCCTGCC 492  
 Db 498 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuPro 517

QY 493 GGCAAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAGGTGCCAGTAC 552  
 Db 518 GlyLysTrpArgProLysMetIleGlyGlyIleGlyGlyPheLysValArgGlnTrp 537

QY 553 GACGAGATCCTGATCGAGATCTCGCGCAAGAAGCCCATCGCACCGTGTGATCGGCC 612  
 Db 538 AspGlnIleProIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyPro 557

QY 613 ACCCCGCGTGAACATCATCGCGCGCAACATGTGATCGCCAGCTGGGTGACCTGAACCTC 672  
 Db 558 ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPhe 577

QY 673 CCCATCAGCCCCCATCGAGACCGTGCCTGAAGCTGAAGCCCGCATGAGCGCCCCCAAG 732  
 Db 578 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 597

QY 733 GTCAAGCAGTGGCCCTGACCCGAGGAGAAGATCAAGGCCCTGACCGCCATCTCGAGGAG 792  
 Db 598 ValLysGlnTrpProLeuThrGluLysIleLysAlaLeuThrAlaIleCysGluGlu 617

QY 793 ATGGAGAAGGAGGCAAGATCAACAAGATCGCGCGCCGAGAACCCCTACACACCCCGTG 852  
 Db 618 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTrpAsnThrProVal 637

QY 853 TTCGCCATCAAGAAGAGGAGGACGACCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTG 912  
 Db 638 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 657

QY 913 AACAAAGCCACCCAGGACTTCTGGGAGTGCAGCTGGGCATCCCCACCCCGCGCGCTG 972  
 Db 658 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 677

QY 973 AAGAAGAAGAAGAGCTGACCGTGTGGAGCTGGGCGAGCGCTACTTACCGCTGCCCTG 1032  
 Db 678 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu 697

QY 1033 GACGAGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACACGAGAGACCCC 1092  
 Db 698 AspLysSerPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrPro 717

QY 1093 GGATCCGCTTACCAGTACAACTGCTGCCCCCAGGGCTGGAAGGCGAGCCCCAGCATCTC 1152  
 Db 718 GlyLeuArgTrpGlnTrpAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 737

QY 1153 CAGAGAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGTGATC 1212  
 Db 738 GlnSerSerMetThrLysIleLeuGluProPheArgAlaLysAsnProGluLeuValIle 757

QY 1213 TACCACTACATGACCACTGTACGTGGCGAGCGACTGGAGATCGCGAGCAGCCGCGCC 1272  
 Db 758 TyrGlnTrpMetAspAspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgAla 777

QY 1273 AAGTCGAGGAGCTGGCGCAAGCACCTGTGCTGGGGCTTACCAACCCCGCAAGAAG 1332  
 Db 778 ArgIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 797

QY 1333 CACGAGAGAGCCCTTCTGTGTGATGGGTACAGCTGCACCCCGCAAGTGGACC 1392  
 Db 798 HisGlnLysGluProProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThr 817

QY 1393 GTCAGCCCATCGAGCTGGCGGAGAGGAGCTGGACCGTGAACGACATCCAGAGAGCTG 1452  
 Db 818 ValGlnSerIleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 837

QY 1453 GTGGGCAAGCTGAAGTGGCGCAGCCAGATCTACCCCGCATCAAGGTGCGCCAGCTGTGC 1512

Db	838	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys	857
Qy	1513	AAGTCTGCTGCGCGCCCAAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGCCGAG	1572
Db	858	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProThrGluGluAlaGlu	877
Qy	1573	CTGAGCTGCGCGAGAACCGGAGATCTCGCGAGCCCGTGACGCGGTGTACTACGAC	1632
Db	878	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp	897
Qy	1633	CCCAGCAGGACCTGCTGCGCGAGATCCAGAGCAGGCGCCACGACCTGACCTACCAG	1692
Db	898	ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspGlnTrpThrTyrGln	917
Qy	1693	ATCTACCAGGAGCCCTTCAAGACCTGAAGACCCGCGCAAGTACGCCAAGATCGCACCGCC	1752
Db	918	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	937
Qy	1753	CACACCAACGACGTGAGCAGCTGACCGAGCCCGTGACAGATCGCCATCGAGAGCATC	1812
Db	938	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle	957
Qy	1813	GTGATCTGGGCGACAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGGGAGCC	1872
Db	958	ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla	977
Qy	1873	TGCTGGACCGACTACTGGCAGCGCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCC	1932
Db	978	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	997
Qy	1933	CCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCGAGACCTTC	1992
Db	998	ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe	1017
Qy	1993	TACGTGACCGCGCGCCCAAGCGGAGACCAAGATCCGCAAGCGCGGTACGTGACCGAC	2052
Db	1018	TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp	1037
Qy	2053	CGGGCGCGGACAGATCGTGAGCTGACCGAGACCCACCAACAGAGACCGAGCTGCAG	2112
Db	1038	LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln	1057
Qy	2113	GCCATCCAGCTGCGCTGACGAGCAGCGGCGAGCGAGTGAAACATCGTGACCGAGCCAG	2172
Db	1058	AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln	1077
Qy	2173	TACGCGCTGGGCATCATCCAGCGCCCGACCAAGAGCGAGCGAGCTGGTGAACCCAG	2232
Db	1078	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln	1097
Qy	2233	ATCATCGAGCAGCTGATCAAGAAGAGAGGTGTACCTGAGCTGGTGGCGCCCAACAAG	2292
Db	1098	IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys	1117
Qy	2293	GGCATCGCGCGCAACGACGACATGCAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTG	2352
Db	1118	GlyIleGlyGlyAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu	1137
Qy	2353	TTCTCGACCGCATCGAT	2370
Db	1138	PheLeuAspGlyIleAsp	1143

Search completed: June 2, 2005, 05:02:46  
Job time : 301.175 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 02:53:39 ; Search time 40.9328 Seconds  
(without alignments)

11607.276 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 4583

Sequence: 1 gtcagccaccatggccga.....gggctagcaccgtgaattc 2469

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-Q=/cgn2.1/USPTO.spool/US09610313/runat.31052005.155137.15159/app.query.fasta.1.7893  
-DB=PIR\_79 -QWTS=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCU=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1.189 @runat.31052005.155137.15159 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR 79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3543.5	77.3	1002	2	S54378 pol polyprotein -
2	3535.5	77.1	1002	1	GNLJND HIV-1 retropepsin
3	3530	77.0	1003	1	GNVWLV HIV-1 retropepsin
4	3513	76.7	1015	1	GNVWH3 HIV-1 retropepsin
5	3491	76.2	1003	2	T09440 pol polyprotein -
6	3488	76.1	1003	1	B44001 HIV-1 retropepsin
7	3487	76.1	1003	1	GNVWA2 HIV-1 retropepsin
8	3477	75.9	1012	1	GNVWYL HIV-1 retropepsin
9	3249	70.9	1027	1	GNLJ51 HIV-1 retropepsin
10	3138	68.5	912	2	S33980 pol polyprotein -
11	3132	68.3	902	2	T01668 pol polyprotein -
12	2809	61.3	559	2	B47175 reverse transcript
13	2789	60.9	559	2	A47175 reverse transcript
14	2388	52.1	1039	2	S46347 pol polyprotein -

15	2244	49.0	1055	2	S53092 pol polyprotein -
16	2240.5	48.9	1055	1	GNLJST HIV-1 retropepsin
17	2231.5	48.7	1056	1	GNLJG3 HIV-1 retropepsin
18	2231	48.7	1036	1	GNLJG2 HIV-1 retropepsin
19	2223	48.5	1034	1	GNLJCA HIV-1 retropepsin
20	2222.5	48.5	1035	1	GNLJGG HIV-1 retropepsin
21	2212.5	48.3	1019	2	T11560 pol polyprotein -
22	2206.5	48.1	1009	2	S28081 pol polyprotein -
23	2185.5	47.7	1032	2	S12153 pol polyprotein -
24	2108	46.0	1058	2	S08436 pol polyprotein -
25	2088	45.6	1061	1	GNLJG4 HIV-1 retropepsin
26	2054.5	44.8	1054	1	GNLJG5 HIV-1 retropepsin
27	1998.5	43.6	656	2	S30484 pol polyprotein -
28	1969.5	43.0	656	2	S30483 pol polyprotein -
29	1428	31.2	1124	2	S23820 pol polyprotein -
30	1425	31.1	1124	1	GNLJFP HIV-1 retropepsin
31	1382	30.2	1124	2	B45557 HIV-1 retropepsin
32	1172	25.6	1146	1	GNLJ22 HIV-1 retropepsin
33	1172	25.6	1146	1	GNLJEW HIV-1 retropepsin
34	1169.5	25.5	1109	1	B45345 HIV-1 retropepsin
35	1152.5	25.1	1087	2	QJ1162 Pol protein - Maed
36	1152.5	25.1	1145	1	GNLJEB HIV-1 retropepsin
37	1149.5	25.1	1086	1	B46335 HIV-1 retropepsin
38	1142	24.9	1101	1	B45390 HIV-1 retropepsin
39	1138	24.8	1101	1	GNLJVS HIV-1 retropepsin
40	1128	24.6	216	2	S43127 reverse transcript
41	1109	24.2	219	2	S32152 RNA-directed DNA p
42	1105	24.1	219	2	S32118 RNA-directed DNA p
43	1105	24.1	219	2	S32139 RNA-directed DNA p
44	1103	24.1	219	2	S32089 RNA-directed DNA p
45	1103	24.1	219	2	S32160 RNA-directed DNA p

ALIGNMENTS

RESULT 1

S54378  
pol polyprotein - human immunodeficiency virus type 1 (fragment)  
C;Species: human immunodeficiency virus type 1, HIV-1  
C;Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
C;Accession: S54378  
R;Theodore, T.; Buckler-White, A.J.  
submitted to the EMBL Data Library, July 1989  
A;Reference number: S54377  
A;Accession: S54378  
A;Status: preliminary  
A;Molecule type: genomic RNA  
A;Residues: 1-1002 <THE>  
A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:G329377; PIDN:AAA5366.1; PID:G3293  
C;Superfamily: pol polyprotein.  
C;Keywords: polyprotein

Alignment Scores:  
Pred. No.: 7.65e-156 Length: 1002  
Score: 3543.50 Matches: 659  
Percent Similarity: 96.25% Conservative: 34  
Best Local Similarity: 91.53% Mismatches: 24  
Query Match: 77.32% Indels: 3  
DB: 2 Gaps: 2

US-09-610-313B-30 (1-2469) x S54378 (1-1002)

Qy	220	TTCTTCGCGGAGGACCTGGCTTCCCGGAGGAGCCCGCGAGTTCCCGAGGAGCAG	279
Db	1	PhePheargGluapLeuAlaPheProGlnGlyalaGlyGluLeuSerSerGluGln	20
Qy	280	AACGCGCCCAACGCCGCCACGCGGAGTGTGAGTGGCGGC	336
Db	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu	40
Qy	337	AGCGAGCGCGCGCGGAGCGGAGCCACCTGACTTC	390
Db	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu	60

```
QY 391 TGGCAGCGCCCTGGTGGACATCAAGGTGGCGGCCAGATCAAGGAGGCCCTGCTGGAC 450
Db 61 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80
QY 451 ACCGGCGCCGACGACACCGCTGCTGGAGGAGATGAGCCCTGCCCGCAAGTGAAGCCCAAG 510
Db 81 ThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100
QY 511 ATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCCAGTACGACCGATCTCGATCGAG 570
Db 101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu 120
QY 571 ATCTGGCGCAAGAGGCCATCGCACCGTCTGATCGGCCCCACCCCGTGAACATCATC 630
Db 121 IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140
QY 631 GCGCGCAACATGTGACCCAGCTGGGTGCGACCCCTGAACCTTCCCATCGACCCCATCGAG 690
Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160
QY 691 ACCGTGCCCCGTGAAGTGAAGCCCGGATGGACGGCCCCCAAGGTGAAGCGATGGCCCCCTG 750
Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180
QY 751 ACCGAGGAGAGATCAAGGCCCTGACCCCATCTGCGAGGAGATGGAGAGGAGGGCAAG 810
Db 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200
QY 811 ATCACCAGATCGGCCCGGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAGAAG 870
Db 201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220
QY 871 GACAGACCAAGTGGCGCAAGCTGGTGGACTTCGGCGAGCTGAACAGCGCCACCCAGGAC 930
Db 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240
QY 931 TTCTGGGAGGTGGAGCTGGGCATCCCCACCCCGCGCTGAAGAGAGAGAGAGCGTG 990
Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260
QY 991 ACCGTGCTGACGTGGCGCGACGCTACTTTCAGCGTGCCTTCGACGAGGACTTCCGCAAG 1050
Db 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280
QY 1051 TACACCGCCTTACCATCCCGACGATCAACAAAGAGACCCCGCGCATCCGCTACCAGTAC 1110
Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300
QY 1111 AAGCTGCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCAGAGCAGCATGACCAAG 1170
Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320
QY 1171 ATCTTGAGCGCTTCCGCGCGCCCAACCCCGAGATCGTGTATACCAAGTACATGGACGAC 1230
Db 321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340
QY 1231 CTGTACGTGGCGCGACCTGGAGATCGGCCAGCACCGCGCCCAAGATCGAGGAGCTGGCG 1290
Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360
QY 1291 AAGCACTGTGCGCTGGCGCTTCCACCCCGCGACCAAGAGCAGCAAGAGGAGCCCGCC 1350
Db 361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380
QY 1351 TTCTGTGGTGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTG 1410
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400
QY 1411 CCGAGAGAGAGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGCAAGCTGAACTGG 1470
Db 401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420
```

```
QY 1471 GCCAGCCAGATCTACCCCGGCATCAAGGTGGCGGCAGCTGTCAAGCTGTGCGCGCGCC 1530
Db 421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuAspGlyThr 440
QY 1531 AAGGCTGTGACCCGACATCGTGGCCCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAAAC 1590
Db 441 LysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGluAsn 460
QY 1591 CGCGAGATCTCTGGCGCGCGCGTGCACGGGTGCTACTACGACCCGACGAGGAGCCTGGTG 1650
Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480
QY 1651 GCCGAGATCCAGAAAGAGGCGCCACGACCTGAGTACCTTACCAGATCTTACCAGAGCCCTTC 1710
Db 481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500
QY 1711 AAGAACTTGAAGACCGGCAAGTACGCAAGATCGGCACCGCCCAACACCAAGCGCTGAAG 1770
Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520
QY 1771 CAGCTGACCGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTATCTGGGCAAGACCC 1830
Db 521 GlnLeuAlaGluValValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540
QY 1831 CCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGTGTGGACCGACTACTGG 1890
Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpValGluTyrTrp 560
QY 1891 CAGGCACTGTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGG 1950
Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580
QY 1951 TACGAGTGGAGAGAGAGCCCATCATCGCGCGGAGACCTTCTACGTGGACCGCGCGCGCC 2010
Db 581 TyrGlnLeuGlnLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600
QY 2011 AACCGGAGACCAAGATCGCAAGCGCGGTACGTGACCGACCGCGCGCGCGCAAGATC 2070
Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620
QY 2071 GTGAGCTGACCGAGACCCAAACCCAGAGACCGAGCTGAGGCCATCCAGCTGGCCCTG 2130
Db 621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
QY 2131 CAGCAGACGGCGCAGCGAGGTGAACATCGTCACCGACGAGCAGTACGCGCTGGGCATCATC 2190
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660
QY 2191 CAGGCGCCGACCGACAGAGCGAGCGAGCTGGTGAACACGAGATCATCGAGCAGCTGATC 2250
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680
QY 2251 AAGAAGAGAAAGTGTACTGTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2310
Db 681 LysLysLysLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700
QY 2311 CAGATCGACAAGCTGTGTGACGAGGCGATCCGGAAGTGTGTCTCTGGACGCGCATCGAT 2370
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
RESULT 2
GNNLJND
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: JQ0067
R;Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.;
Gene 81, 275-284, 1989
A;Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunoc
A;Reference number: JQ0065; MUID:90034200; PMID:2806917
A;Accession: JQ0067
```

A:Molecule type: DNA  
A:Residues: 1-1002 <SPI>  
A:Cross-references: UNIPROT:P18802; GB:M27323; NID:G328154; PID:AAA44869.1; PID:G328158  
C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
C:Genetics:  
A:Gene: pol  
C:Superfamily: pol polyprotein  
C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
F:56-154/Product: retropepsin #status predicted <RTP>  
F:80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
Pred. No.: 1,79e-155 Length: 1002  
Score: 3535.50 Matches: 659  
Percent Similarity: 96.25% Conservative: 34  
Best Local Similarity: 91.53% Mismatches: 24  
Query Match: 77.14% Indels: 3  
DB: 1 Gaps: 2

US-09-610-313B-30 (1-2469) x GNLJND (1-1002)

Qy	220	TTCTTCGCGAGGACCTGGCCCTTCCCGAGGCAAGCCCGCGAGTTCCTCCCGAGCGAGCAG	279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyysAlaGlyGluPheSerSerGluGln	20
Qy	280	AACCGCCCAACACGCCCCACGCGCGAGCTGCGGTG---CGCGCGCAACCCCGCGC	336
Db	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu	40
Qy	337	AGCGAGCGCGCGCGAGCGCGCGAGCCCTG-----ACCTCCCGCGAGTTCCTCCCG	390
Db	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnThrLeu	60
Qy	391	TGGCAGCGCCCGCTGGTGAGCATCAAGTGGCGCGCGAGTCAAGAGGCGCCCTCTCGAC	450
Db	61	TrpGlnArgProLeuValThrIleIleIleIleIleIleIleIleIleIleIleIleIleIle	80
Qy	451	ACCGCGCCGACGACACCGTCTGGAGAGATGAGCTGCCCGCAAGTGGAGCCCAAG	510
Db	81	ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys	100
Qy	511	ATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGAGTACGACCGAGTCTGATCGAG	570
Db	101	MetIleGlyIleIleGlyIlePheIleLysValArgGlnTrpAspGlnIleLeuIleGlu	120
Qy	571	ATCTGCGCAAGAGGCATCGGCACCGTGTGATCGCGCGCGAGTCAAGTGGAGCATCATC	630
Db	121	IleCysGlyTyrlsAlaMetGlyThrValLeuValGlyProThrProValAsnIleIle	140
Qy	631	GGCGCAACATGTGACCCAGCTGGGCTGCACCTGAACTTCCCATCAGCCCGCATCGAG	690
Db	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
Qy	691	ACCGTCCCGTGAAGTGAAGCCCGGATGACCGCGCGCGAGTGAAGTGAAGTGGCCCTG	750
Db	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
Qy	751	ACCGAGGAGAAGTCAAGGCCCTGACCGCATCTCGCGAGGAGTGAAGAGGGGCAAG	810
Db	181	ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys	200
Qy	811	ATCACCAAGATCGCGCGCGAGAACCCCTACAACACCCCGCTGTTCGCGCATCAAGAGAAG	870
Db	201	IleSerArgIleGlyProGluAsnProTyArgThrPheAlaIleLysLysLys	220
Qy	871	GACGACCAAGTGGCGCGAGCTGGTGGTCTTCGCGAGCTGAACAGCGCGACCCAGGAC	930
Db	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
Qy	931	TTCTGGGAGTGGAGCTGGGCATCCCCACCCCGCGCGCTGAAGAGAAGAGAGCGTG	990
Db	241	PhetrpGluValGlnLeuGlyIleProHieProAlaGlyLeuLysLysLysLysServAl	260

Qy	991	ACCGTGTGAGCTGGGCGAGCCCTACTTACGCTGCCCTCGAGCGAGGACTTCGCGAAG	1050
Db	261	ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArgLys	280
Qy	1051	TACACCCCTTTACCATCCCGCAGCATCAACAACAGACCCCGCGCATCCGCTACCGATAC	1110
Db	281	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrlsGlnTy	300
Qy	1111	AACGTGTGCGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTTCAGAGCAGCATGACCAAG	1170
Db	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
Qy	1171	ATCCTGAGCGCTTCCCGCGCGCGCAACCCCGAGATCTGATCTACCATGATGACACAC	1230
Db	321	IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrlsGlnTyrlsMetAsp	340
Qy	1231	CTGTACGTGGCGAGCGACCTGGAGATCGCGCAGCAGCGCGCAAGATCGAGAGCTGGCG	1290
Db	341	LeuTyrlsValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	360
Qy	1291	AAGCACCTGTCTGGCTGGGCTTCCACACCCCGCGAGCAAGAGACCAAGAGGAGCGCC	1350
Db	361	GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	380
Qy	1351	TTCTGTGGATGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAGCTG	1410
Db	381	PhetrpMetGlyTyrlsGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu	400
Qy	1411	CCCGAGAGGAGAGCTGGACCGTGAACGAGCATCCAGAGCTGTGCAAGCTCTCGCGCG	1470
Db	401	ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	420
Qy	1471	GCACGCGAGATCTACCCCGCGCATCAAGTGTGCGCGAGCTGTGCAAGCTCTCGCGCG	1530
Db	421	AlaSerGlnIleTyrlsAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr	440
Qy	1531	AAGCCCTGACCGACATCTGCTGCGCCCTGACCGAGAGCGCGAGCTGGAGCTGGCGAG	1590
Db	441	LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn	460
Qy	1591	CGCGAGATCTCGCGAGCGCGTGCACGGCTGTACTACGACCCCGAGCAAGCAAGCTGTG	1650
Db	461	ArgGluIleLysLysGluProValHisGlyValTyrlsAspProSerLysAspLeuIle	480
Qy	1651	GCGGAGATCCAGAGCAGGCGCGACGACGAGTGGACCTTACGAGATCTACGAGCGCCCTT	1710
Db	481	AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrlsGlnIleTyrlsGlnProPhe	500
Qy	1711	AAGAACCTGAAGACCGCGCAAGTACGCCAAGATGCGCACCGCGCCACACCAACGAGCTG	1770
Db	501	LysAsnLeuLysThrGlyLysTyrlsAlaArgThrArgGlyAlaHisThrAsnAspValLys	520
Qy	1771	CAGTGCACGAGCGCGTGCAGAGATCGCCATGAGAGCATCTGTGATCTGGGCGCAAGAC	1830
Db	521	GlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr	540
Qy	1831	CCCAAGTTCGCGTCCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG	1890
Db	541	ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrlsTrp	560
Qy	1891	CAGGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGGTGG	1950
Db	561	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp	580
Qy	1951	TACGAGTGCAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC	2010
Db	581	TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrlsValAspGlyAlaAla	600
Qy	2011	AACCGCGAGACCAAGATCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2070
Db	601	AsnArgGluThrLysLeuGlyLysAlaGlyTyrlsValThrAspArgGlyArgGlnLysVal	620
Qy	2071	GTGAGCCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGCGCCATCTCAGTGGCCCTG	2130



```
Db 621 valpropheThrAspThrThrAnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 2131 CAGGACAGCGCGAGCGAGGTGAACATCGTACCCAGCAGCGAGTACGCGCTTGGGCATCATC 2190
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 2191 CAGGCCAGCGCCCAACAAGACGAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCGAGCTGATC 2250
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 2251 AAGAAGGAGAGGTGTACCTGAGCTGGCTCCCGCCACAAAGGCGCATCGCGCGCAACGAG 2310
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 681 LysLysGlnLysValTyrLeuAlaIleValProAlaHisLysGlyIleGlyIleGlnGlu 700
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 2311 CAGATCGACAAAGTGTGTGACAAAGGCGATCCGCAAGGTGTGTCTTGGACGCGCATCGAT 2370
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

RESULT 3
GNVWL
N;Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C;Species: human immunodeficiency virus type 1, HIV-1
A;Note: host Homo sapiens (man)
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002
C;Accession: A03966
R;Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A;Title: Nucleotide sequence of the AIDS virus, LAV.
A;Reference number: A90866; MUID:85099333; PMID:2981635
A;Accession: A03966
A;Molecule type: DNA
A;Residues: 1-1003 <WAI>
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
A;Gene: pol
C;Superfamily: pol polypeptide
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle
F;57-155/Product: retropepsin #status predicted <RTP>
F;81/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:
Pred. No.: 3,2e-155 Length: 1003
Score: 3530.00 Matches: 659
Percent Similarity: 95.98% Conservative: 33
Best Local Similarity: 91.40% Mismatches: 25
Query Match: 77.02% Indels: 4
DB: 1 Gaps: 2

US-09-610-313B-30 (1-2469) x GNVWL (1-1003)
QY 220 TTCTTCGCGAGGACCTGGCGCTTCCCGCAGGCAAGCGCGCGAGTTCCCGCAGCGAGCAG 279
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyIleAlaArgGluPheSerSerGluGln 20
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 280 AACCGGCCCAACAGCCCAACAGCGCGAGCTGAGGTG-----CGCGCGCAACACCC 333
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTyrGlyArgAspAsnSer 40
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 334 CGCAGCGAGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 388 CTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 61 LeuTyrGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 448 GACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 81 AspThrGlyAlaAspAspThrValLeuGluMetSerLeuProGlyArgTyrLysPro 100
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
QY 508 AAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
```

```
101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
568 GAGATCTGGCGCAAGAGCCATCGGCACCGCTGCTGATCGGCCCCACCCCGCTGAACATC 627
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
628 ATCGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACCTCCCATCAGCCCATC 687
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
688 GACACCGTGCCTGAGCTGAAGCCCGCGCATCGAGCGCGCCCAAGGTGAAGCAAGTGGCC 747
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTyrPro 180
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
748 CTGACCGAGAGAGATCAAGCCCTGACCGCCCATCTCTGAGGAGAGATGAGAGAGAGGCG 807
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
808 AGATCACCAAGATCGCGCGCGAGAACCCCTACAACACACCCCGCTTTCGCATCAAGAAG 867
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
868 AAGCAGACACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAG 927
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
221 LysAspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
928 GACTTCTGGAGGTGAGCTGGGCATCTCCACCCCGCGCGCTGAGAGAGAGAGAGAGAGC 987
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
241 AspPheTyrGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
988 CTGACCGTGTGAGCTGGCGGAGCGCTACTTCCAGGTGCGCCCTGAGCAGGAGCTCCGC 1047
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1048 AAGTACACCGCTTCCACCTCCCGAGCATCAACAAGAGAGACCCCGCGCATCCGCTACAG 1107
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1108 TACAAGTGTGCTGCCAGGCTGGAAGGAGCGAGCCGAGCATCTTCAGAGCAGCATGACC 1167
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
301 TyrAsnValLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnSerSerMetThr 320
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1168 AAGATCTCGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1227
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
321 LysIleLeuGlnProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1228 GACCTGTAGCTGGCGCAGCGACCTGGAGATCGGCAGCACCGCGCCCAAGATCGAGGAGCTG 1287
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1288 CGCAAGACCTGTGCTGGCGCTTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1347
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
361 ArgGlnHisLeuLeuArgTyrGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1348 CCTTCTCTGTGAGTGGCTACGAGCTGCACCGCGAGTGGACCGCGCGCGCGCGCGCGCGCG 1407
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
381 ProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThrValGlnProIleVal 400
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1408 CTCCCGCAGAGAGAGAGCTGGACCGCTGAACACGATCCAGAAAGCTGTGGGCAAGCTGAAC 1467
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
401 LeuProGluLysAspSerTyrThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1468 TGGGCGAGCGCAGATCTACCCCGCGCATCAAGGTGGCGCGAGCTGTGCAAGCTGTGCGCGCG 1527
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1528 GCCAAGCGCTTGACCGACATCGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGCGAG 1587
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1588 AACCGGAGATCTCTGGCGGAGCGCGCTGACGGGTGTACTACGACCCCGAGAGGAGCTGTG 1647
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
```



```
Db 281 AlatyPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
|||
QY 1072 AGCATCAACACGAGACCCCGCATCCGCTACCACTGACAACTGCTGCCCGCAGGCTGG 1131
|||
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
|||
QY 1132 AAGGGCAGCCCGCAGACTCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
|||
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
|||
QY 1192 CGCAACCCCGAGATCGTGATCTACAGTACATGACGACCTGTACGTGGCAGCGACCTG 1251
|||
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
|||
QY 1252 GAGATCGCCGAGCAGCCGCGCAAGATCGAGAGCTGGCAGCACCTGCTGCCCTGGGC 1311
|||
Db 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
|||
QY 1312 TTCACCAACCCCGACAGAGCAGCAGAGAGCCCTTCTGTGGATGGGTACGAG 1371
|||
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
|||
QY 1372 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCCGAGAGGAGAGCTGGACC 1431
|||
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
|||
QY 1432 GTGAACGACATCCAGAAGCTGTGGCAAGCTGAATGGCGCCAGCCAGATCAACCCGCGC 1491
|||
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
|||
QY 1492 ATCAAGTGTGCCAGCTGTGCAAGCTGCTCGCGCGCCCAAGCCCTACCGACATCGTG 1551
|||
Db 441 IleIysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
|||
QY 1552 CCCTGTACCGAGAGGCGCAGCTGGAGCTGCCGAGAACCGCGAGATCTGTGCGGAGCCC 1611
|||
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
|||
QY 1612 GTGCACGGCTGTACTACGACCCCGCAGCAAGACTGTGTGGCCGAGATCCAGAGCAGGCG 1671
|||
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
|||
QY 1672 CACGACAGTGGACTACACAGATCTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1731
|||
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
|||
QY 1732 TACGCCAAGATGCGCACCCGCCACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAG 1791
|||
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
|||
QY 1792 AGATCCGCATGGAGACATCGTGATCTGGGGCAGACCCCAAGTTCGCCCTGCCCATC 1851
|||
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
|||
QY 1852 CAGAGGAGACCTGGGAGACTGTGTGACCGACTACTTGGCAGGCGCACTGATCCCGAG 1911
|||
Db 561 GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
|||
QY 1912 TGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGGTACCACTGGAGAGGAGGCC 1971
|||
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuLeuTrpTyrGlnLeuGluLysGluPro 600
|||
QY 1972 ATCATCGCGCGGAGACTTCTACGTGAGCGCGCGCCCAACCGCGAGACAAGATCGGC 2031
|||
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
|||
QY 2032 AAGCCGGCTACGTGACCGACCGGGCGCGCAGAAAGATCTGAGCCCTGACCGACCAACC 2091
|||
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
|||
QY 2092 AACCAAGAAGCCGAGCTGCAGGCGCATCCAGCTGGCCCTCGAGGACACCGCGCAGGAGTG 2151
|||
```

```
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
|||
QY 2152 AACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGACCAAGAGC 2211
|||
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
|||
QY 2212 GAGACGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTG 2271
|||
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
|||
QY 2272 AGCTGGTGCCCGCCACACAGGCGATCGCGGCAACGACGATCCACAGCTGTGAGC 2331
|||
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlnGluGlnValAspLysLeuValSer 720
|||
QY 2332 AAGGCGATCCGCAAGTGTCTTCTTCGACGCGCATCGAT 2370
|||
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
|||
RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRFL) (fragment)
C:Species: human immunodeficiency virus type 1, HIV-1
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T09440
R:Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A:Reference number: Z16673
A:Accession: T09440
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1003 <PAN>
A:Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C:Genetics:
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 1.99e-153 Length: 1003
Score: 3491.00 Matches: 650
Percent Similarity: 95.56% Conservative: 39
Best Local Similarity: 90.15% Mismatches: 28
Query Match: 76.17% Indels: 4
DB: 2 Gaps: 2

US-09-610-313B-30 (1-2469) x T09440 (1-1003)
QY 220 TTCTTTCGGCAGGACTGGCCCTTCCCGCAGGCAAGCCCGAGTTCCCGCAGGAGCAG 279
|||
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
|||
QY 280 AACCGCGCAACAGCCCGCCAGCCGCGAGCTCGAGTG-----CGCGCGCAACACCC 333
|||
Db 21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40
|||
QY 334 CGCAGGCGAGCCCGCGCCGAGCGCCAGGGC-----ACCCTGAACCTTCCCGCAGATCAC 387
|||
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60
|||
QY 388 CTGTGGCAGCGCCCGCTGTGACCATCAAGTGGCGGCGCAGATCAGGAGGCGCCCTGCTG 447
|||
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
|||
QY 448 GACACCGCGCGCCAGCACACCGTGTGGAGGAGATGAGCTGCGCGCAAGTGAAGCC 507
|||
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
|||
QY 508 AAGATGATCGCGCGCATCGCGCTTCATCAAGGTGCGCCAGTACGACCGATCCTGATC 567
|||
Db 101 LysMetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
|||
QY 568 GAGATCTGGCGGCAAGAGCGCATCGCCACTGCTGATCGGCCCGCCACCCCGTGAACATC 627
|||
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyThrValLeuValGlyProValAsnIle 140
|||
```

QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGCTGCAACCTGAACTTCCCATCAGCCCCATC 687  
Db 141 IleglyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
QY 688 GAGACCTGCGCGTGAAGCTGAAGCCCGGATGACGCGCCCAAGGTGAAGCAGTGGCCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGAGGAGAAAGATCAAGCGCCCTCACCGCCATCTCGAGGAGATGAGGAAGAGGCG 807  
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AAGATCACCAAGATCGCGCCCGGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAAG 867  
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
QY 868 AAGGACAGCAACCAAGTGGCGCAAGCTGTGACCTTCGCGAGCTGAACAGCGCACCCAG 927  
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysLysThrGln 240  
QY 928 GACTTCTGGAGGTGACGTGGGATCCCGGATCCCGCCCGGCTGAAGAAGAAAGAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260  
QY 988 GTGACCTGTGACGTGGGCGACGCTTACTTACGCTGCCCTGGACGAGGACTTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCCCTTACCATCCCGACGATCAACAAACGAGACCCCGGATCCGCTACCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGlnThrProGlyIleArgTyrGln 300  
QY 1108 TACAACCTGTGCCCCCAGGGGTGAAGGCGAGCCCGGATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATGTGATCTACGATCATCGAC 1227  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340  
QY 1228 GACTGTACCTGGGCGGACCTGGAGATGGCGGACGCGGACCGGCGACAGATCGAGGACTG 1287  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360  
QY 1288 CGCAAGCACCTGCTGCGCTGGGCTTCAACACCCCGGACAGAGACCAAGAGAGGAGCC 1347  
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1348 CCCTTCTGTGGATGGCTACGAGCTGCACCCGACAAAGTGGACCGTGCAGCCCATCGAG 1407  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1408 CTGCCCCGAGAGGAGACTGGACCGTGAACGACATCCAGAGCTGGTGGCGAGCTGAAC 1467  
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1468 TGGCGCAGCAGATCTACCCGCGCATCAAGTGGCGGCGAGCTGTGCAAGCTGCTGCGCGGC 1527  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440  
QY 1528 GCCAAGCCCTTGACCGCATGTGTGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAG 1587  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1588 AACCGGAGATCTGCGCGAGCCCGTGCACGGGTGTACTACGACCCCGGACGAGGACTG 1647  
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480  
QY 1648 GTGCGCGAGATCCAGAGCAGGCGCACGACAGTGGACCTACAGATCTACCGAGGAGCC 1707  
Db 481 IleAlaGluLeuGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500

QY 1708 TTCAGGAACCTGAGAGCCGCGAAGTACGCCAAGATGCGCACCGCCACCAACGAGCTG 1767  
Db 501 PheLysIleLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520  
QY 1768 AAGCAGCTGACCGAGGCGGTGCAGAAGATGCCCATCGAGAGCATCGTGATCTGGGGCAAG 1827  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540  
QY 1828 ACCCCCAAGTTCGCGCTGCCATCCAGAAAGAGACCTGGGAGACCTGGTGACCGACTAC 1887  
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560  
QY 1888 TGGCAGGCGACCTGGATCCCGGAGTGGAGCTTCGTGAACACCCCGCTGTGAAGCTG 1947  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
QY 1948 TGGTACCAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAGCGGCGCC 2007  
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
QY 2008 GCCAACCGCAGACCAAGATCGCAAGCGCGCTTACGTGACCGCGGCGCGCGCAAG 2067  
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620  
QY 2068 ATCGTGACCTGACCGAGACCAACCAAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2127  
Db 621 ValValSerLeuThrAspThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
QY 2128 CTGCAAGACAGCGCAGGAGTGAACATCGTACCGACAGCAGCAGTAGTAGCGCTGGGATC 2187  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
QY 2188 ATCCAGCGCCAGCCGCAAGAGCAGAGCTGGTGAACCCAGATCATCGACGAGCTG 2247  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
QY 2248 ATCAAGAGAGAGAGTGTACCTGAGCTGGTGGCGCCGCCCAAGGCGCATCGCGGCAAC 2307  
Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
QY 2308 GAGCAGATCCACAAGCTGGTGAGCAGGGCATCCGCAAGGTGTCTTCTGGAGCGGATC 2367  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2368 GAT 2370  
Db 721 Asp 721  
RESULT 6  
B44001  
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)  
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
C;Species: human immunodeficiency virus type 1, HIV-1  
A;Note: host Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C;Accession: B44001  
J. Virol. 66, 6587-6600, 1992  
A;Title: Complete nucleotide sequence, genome organization, and biological properties of  
A;Reference number: A44001; MUID:93021387; PMID:1404605  
A;Accession: B44001  
A;Molecule type: DNA  
A;Residues: 1-1003 <LII>  
A;Cross-references: UNIPROT:P35963; GB:M93258  
C;Comment: This protein is synthesized as a gag-pol polyprotein.  
C;Comment: Specific enzymatic cleavages may yield mature proteins including proteinase, f  
C;Genetics:  
A;Gene: pol  
C;Superfamily: pol polyprotein  
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo  
F;57-155/Product: retropepsin #status predicted <RTP>  
F;81/Active site: Asp (shared with dimeric partner) #status predicted

## Alignment Scores:

Pred. No.: 2,74e-153 Length: 1003  
Score: 3488.00 Matches: 651  
Percent Similarity: 95.42% Conservative: 37  
Best Local Similarity: 90.29% Mismatches: 29  
Query Match: 76.11% Indels: 4  
DB: 1 Gaps: 2

US-09-610-313B-30 (1-2469) x B44001 (1-1003)

Qy	220	TTCTTCGCGAGGACCTGGCTTCCTCCCGGAGGAGCCCGCGAGTTCCCGAGGAGCAG	279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyLeuAlaArgLysPheSerSerGluGln	20
Qy	280	AACCGCGCCAAACGCCCCACCAGCCGCGAGCTGCAGGTG-----CGCGCGCAACCC	333
Db	21	ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer	40
Qy	334	CGCAGCGAGCGCCGCGCGAGCGCCAGGCGCACCTGTG-----AACTTCCCGCCAGATCAC	387
Db	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr	60
Qy	388	CTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGCGCGCCAGATCAAGGAGGCGCTGTG	447
Db	61	LeuTrpGlnArgProLeuValThrIleLeuIleGlyGlnLeuLysGluAlaLeuLeu	80
Qy	448	GACACGCGCGCCGACGACCGCTGTGGAGGAGATGAGCTGCGCGCAAGTGGAGGCC	507
Db	81	AspThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro	100
Qy	508	AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCAGTACGACACGATCCTGATC	567
Db	101	LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle	120
Qy	568	GAGATCTGCGGCAAGAGCGCATCGGCACCGCTGTGTGATCGGCGCCCGCCCGTGAACATC	627
Db	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
Qy	628	ATCGCGCGACATGTGACCCAGCTGGCTGGCTGACCCCTGACCTGAACCTCCCATCAGCCCATC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
Qy	688	GAGACCTGCGCGTGAAGCTGAAGCCCGCATGACCGCGCCCAAGGTGAAGCATGTGCGCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
Qy	748	CTGACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAGATGAGAGAGGCG	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
Qy	808	AAGATCACCAAGATCGGCGCGGAGAACCCCTACACACCCCGCTGTTCGCCATCAAGAAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys	220
Qy	868	AAGGACACCAAGTGGCGCAAGCTGGTGGATCTTCGCGAGCTGAAACAAGCGCACCCAG	927
Db	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
Qy	928	GACTTCTGGAGGTGACGTGGGATCCCGACCCCGCGCGCTGGAAGAGAGAGAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
Qy	988	GTGACCGTGTGGAGTGGGCGCGCTTCTACGCTGCGCTTCGACGAGGAGCTTCGCG	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuHisGluAspPheArg	280
Qy	1048	AAGTACACCGCTTCACCATCCCGAGCATCAACAACGAGACCCCGCGCATCCGCTACCA	1107
Db	281	LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTyGln	300
Qy	1108	TACACGTGTGCGCGCGTGAAGGCGAGCGCCAGCATCTTCGACGAGCAGCATGACC	1167
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320

Qy	1168	AAGATCCTGGAGCCCTTCGCGCGCCCGCAACCCCGAGATCGTGTATCTACAGTACATGGAC	1227
Db	321	ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTyGlnTyMetAsp	340
Qy	1228	GACCTGTACGTGGGAGCGACCTGGAGATCGCGCACACCGCGCCCAAGATCGAGGAGCTG	1287
Db	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
Qy	1288	CGCAAGCACCTGCTGCTGGGCTTACCAACCCCGCAAGAACAGCAGAGAGGAGCC	1347
Db	361	ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
Qy	1348	CCCTTCTGTGATGGGCTACGAGCTGCACCCCGCAAGTGGACCGCTGCGCCCATCGAG	1407
Db	381	ProPheLeuTrpMetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleVal	400
Qy	1408	CTGCCCGAAGAGGAGCTGGACCGGTGAACGACATCAGAGCTGGTGGGCAAGCTGAAC	1467
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
Qy	1468	TGGCGCAGCCAGATCTACCCCGCGCATCAAGGTGCGCGCAGCTGTCAAGCTGCTGCGCGG	1527
Db	421	TrpAlaSerGlnIleTyThrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly	440
Qy	1528	GCCAAGCGCTGACCCGACATCGTGCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAG	1587
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGlu	460
Qy	1588	AACCGGAGATCTGCGCGAGCGCGTGCACGGGTGTACTACGACCCCGAGCAAGAGCTG	1647
Db	461	AsnArgGluIleLeuLysGluProValHisGlyValTyThrAspProSerLysAspLeu	480
Qy	1648	GTGGCGAGATCCAGAGGAGCGCCACGACCGAGTGCCTTACAGATCTTACAGAGGAGCC	1707
Db	481	IleAlaGluIleGlnLysGlnGlyGlnTrpThrTrpThrGlnIleTyThrGlnLysPro	500
Qy	1708	TTCAAGAACCTGAAAGCCGCAAGTACGCAACGACCGCACCCGCCACACCAACGAGCTG	1767
Db	501	PheLysAsnLeuLysThrGlyLysTyThrAlaArgThrArgGlyAlaHisThrAsnAspVal	520
Qy	1768	AAGCAGCTGACCGAGCGCTGCAAGATGCGCATCGAGAGCATCGTGTATCTGGGGCAAG	1827
Db	521	LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys	540
Qy	1828	ACCCCAAGTTCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTAC	1887
Db	541	ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrTrpThrGluTy	560
Qy	1888	TGGCAGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCGCCCTGTGTGAAGCTG	1947
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	580
Qy	1948	TGTTACAGCTGGAAGAGGCGCCATCTCGCGCGCGAGACCTTCTACGTGGAGCGGCGCC	2007
Db	581	TrpTyThrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyValAspGlyAla	600
Qy	2008	GCCAACCGGAGACCAAGATCGGAGCGCGCTGTGACCGACCGCGGCGCGCGCAAG	2067
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyThrValThrAsnLysGlyArgGlnLys	620
Qy	2068	ATCGTGTGAGCTGACCGAGACCAACACCAAGAGACCGAGCTGCGAGCGCATCCAGCTG	2127
Db	621	ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyThrLeuAla	640
Qy	2128	CTGAGAGACAGCGCGAGGAGTGAACATCGTGCACGACGAGCGAGTACCGCTGGGCGATC	2187
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIle	660
Qy	2188	ATCCAGCGCGCGCGAGGAGCGAGCTGTGTGAACCGAGATCATCGAGCAGCTG	2247
Db	661	IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu	680

QY 2248 ATCAAGAGGAGNAGGTCTACCTGAGCTGGGTGCGCCGCCCAAGAGGATCGGGCGCAAC 2307  
Db 681 lilelyslsGluLysValTyrLeuAlaTrpValProAlaHisLysGlyGlyGlyAsn 700  
QY 2308 GAGCAGATCGCAAGCTGGTGAAGAGGATCGGCAAGGTGCTTCTCGACGGCATC 2367  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2368 GAT 2370  
Db 721 Asp 721  
RESULT 7  
GNVW2  
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate ARV-2)  
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
C;Species: human immunodeficiency virus type 1, HIV-1  
A;Note: host Homo sapiens (man)  
C;Date: 17-May-1985 #sequence\_revision 17-May-1985 #text\_change 09-Jul-2004  
C;Accession: A03968  
R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-Sh  
Science 227, 484-492, 1985  
A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).  
A;Reference number: A04003; MUID:85090453; PMID:2578227  
A;Accession: A03968  
A;Molecule type: DNA  
A;Residues: 1-1003 <SAN>  
C;Cross-References: UNIPROT:P03369; GB:K02007; NID:g328658; PTDN:AAB59876.1; PID:g328662  
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
C;Genetics:  
A;Gene: pol  
C;Superfamily: pol polyprotein  
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
F;57-147/Product: retropepsin #status predicted <RTP>  
F;81/Active site: Asp (shared with dimeric partner) #status predicted  
Alignment Scores:  
Pred. No.: 3,04e-153 Length: 1003  
Score: 3487.00 Matches: 650  
Percent Similarity: 95.42% Conservative: 38  
Best Local Similarity: 90.15% Mismatches: 29  
Query Match: 76.09% Indels: 4  
DB: 1 Gaps: 2  
US-09-610-313B-30 (1-2469) x GNVW2 (1-1003)  
QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCGCGAGTTCCCGAGGACGAG 279  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 280 AACCGCGCCCAACAGCCCGCCAGCGGAGTGCAGGTGCGCGGC-----GACAACCCC 333  
Db 21 ThrArgAlaAsnSerProThrArgGluLeuGlnValTrpGlyGlyGluAsnAsnSer 40  
QY 334 CGAGCCAGGCGCGCGGAGCGCCAGGCGACCTG-----ACTTCCCGCAGATCAC 387  
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
QY 388 CTGTGGCAGCGCCCTGGTGAGCATCAAGGTGGCGGCGCAGATCAAGGAGGCCCTGCTG 447  
Db 61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlyGlnLeuLysGluAlaLeuLeu 80  
QY 448 GACACCGCGCGCAGCACCGCTGCTGGAGGAGATGAGCTGCGCGGCGCAAGTGAAGCCC 507  
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100  
QY 508 AAGATGATCGGCGGATCGGGCTTCATCAAGGTGGCGCAGTACGACGACGATCCTGATC 567  
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProVal 120  
QY 568 GAGATCTGCGCGCAGAGGCCATCGGCACCGTGTGATCGGCCCGCCCGCGTGAACATC 627  
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

QY 628 ATCGGCGCCCAACATGCTGAGCCAGCTGGGTGACACCTTGAACTTCCCATCAGCCCATC 687  
Db 141 lileglyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
QY 688 GAGACCTGCGCGTGAAGCTGAAGCCCGGATGACGCGCCCAAGGTGAAGCAGTGCGCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGAGGAGAAAGATCAAGGCCCTGACCCGCCATCTGCGAGGAGATGGAAGAGGAG 807  
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AAGATCAACAGATCGGCGCGCGAGAACCCCTACAAACACCCCGTGTTCGCCATCAAGAG 867  
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
QY 868 AAGACAGACCAAGTGGCGCAAGCTGGTGACCTTCGCGAGCTGAAACAAGCAGCCAG 927  
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
QY 928 GACTTCTGGGAGTGCAGCTGGGCGATCCCCACCCCGCGCTGAAGAAAGAAAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
QY 988 GTGACCGTGTGAGCTGGGCGACCCCTACTTCAGCGTGCCTCGGAGCAGGACTTCGCG 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
QY 1048 AAGTACACCCCTTCACCATCCCGCAGCATCAACACGAGACCCCGCGCATCCGCTACCG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
QY 1108 TACACCTGTGCGCCCGCAGGCTGAAGGCGAGCCCGCAGCATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTGTGAGCCTTCCGCGCGCCCAACCCCGAGATCGTGATCTACCATCATGAGC 1227  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
QY 1228 GACTGTACCTGGCGGAGCTGGAGATCGGCGAGCAGCCGCGCGCAAGATCGAGGAGCTG 1287  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1288 CGCAAGCACCTGTGCTGCGCTGGGCTTCAACACCCCGCAGAACAGACCAAGAGGAGGCC 1347  
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1348 CCTTCTGTGGATGGGCTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG 1407  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleMet 400  
QY 1408 CTGCGCGAGAGGAGAGCTGGAACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1467  
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1468 TGGCGCAGCAGATCTACTACCCCGCGCATCAAGGTGGCGCGAGCTGTCAAGCTGCTCGCGGC 1527  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleValLysValLysGlnLeuCysLysLeuLeuArgGly 440  
QY 1528 GCCAAGCCCTGACCGCATCGTGGCCCTGACCCCGAGAGGCGCGAGCTGGAGCTGCGCGAG 1587  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1588 AACCGCGAGATCTGCGCGAGCCCGCTGCACCGCGGTGTACTACGACCCCGCAGCAAGACCTG 1647  
Db 461 AsnArgGluIleLeuLysGluProValHisGluValTyrTyrAspProSerLysAspLeu 480  
QY 1648 GTGCGCGAGATCCAGAGCAGGCGCCACGACGAGTGGACCTACACGATCTTACCGAGGCC 1707  
Db 481 ValAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrThrGlnIleTyrGlnGluPro 500





Db 324 PheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343  
Qy 1210 ATCTACCATGATGAGACCTGTGTGGCGGACGACCTGTGGAGATCGGCGCACACCGC 1269  
Db 344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363  
Qy 1270 GCCAAGATCGAGGAGTCGCGAAGACCTGTGTGGCGGCTTCACACCCCGACAG 1329  
Db 364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLys 383  
Qy 1330 AAGCACCAGAGGAGCCCTCTCTGTGGATGGCTACGAGCTGACCCCGACAAGTGG 1389  
Db 384 LysHisGlnLysGluProPheLeuThrMetGlyTyrGluLeuHisProAspLysTrp 403  
Qy 1390 ACCGTGCGCCATCGAGCTCGCGAGAGGAGCTGGACCGTGAACGACATCCAGAAG 1449  
Db 404 ThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLys 423  
Qy 1450 CTGTGGGCAAGCTGAATCTGGGCGAGCGACATCTACCCCGGCATCAAGTGGCGCAGTG 1509  
Db 424 LeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443  
Qy 1510 TGCAAGCTGTGCGCGCGCCCAAGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCC 1569  
Db 444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463  
Qy 1570 GAGCTGAGCTGGCGGAGAACCGGAGATCTCGCGGAGCCCGTGCACGGCGTGTACTAC 1629  
Db 464 GluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyr 483  
Qy 1630 GACCCAGCAAGACCTGTGTGGCGAGATCCAGAAAGCAGGCGCACGACCTGACCTAC 1689  
Db 484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrpThrTyr 503  
Qy 1690 CAGATCTACGAGGAGCCTTCAAGACCTGGAAGCGCGAGTACGCAAGATCGGCACC 1749  
Db 504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523  
Qy 1750 GCCCACCAACGACGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATCGAGAGC 1809  
Db 524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543  
Qy 1810 ATCTGTATCTGGGCAAGACCCCAAGTTCGCTGCGCCATCCAGAGGAGACCTGGAG 1869  
Db 544 IleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGlu 563  
Qy 1870 ACCTGGTGGACGACTGTCAGCGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACC 1929  
Db 564 ThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThr 583  
Qy 1930 CCCCCTCTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCCGAGACC 1989  
Db 584 ProProLeuValLysLeuTyrTyrGlnLeuGlnLysGluProIleValGlyAlaGluThr 603  
Qy 1990 TTCTACGTGGACGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACC 2049  
Db 604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623  
Qy 2050 GACCGGCGCGGAGAGATCGTGAAGCTGACCGAGACCCACCAAGACCGAGACCGAGCTG 2109  
Db 624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643  
Qy 2110 CAGGCCATCCAGCTGGCGCTCGAGCAGCGCGCAGCGAGTGAACATCTGTGACCGCAGC 2169  
Db 644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAlaIleValThrAspSer 663  
Qy 2170 CAGTACCGCTGGGCATCATCAGCGCCCGACCGCAAGAGCAGAGCGAGCTGGTGAAC 2229  
Db 664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683  
Qy 2230 CAGATCATCGAGCAGCTGATCAAGAGGAGAGTGTACTGAGCTGGTGGCCCGCCAC 2289

Db 684 GlnIleIleGluGlnLeuLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHis 703  
Qy 2290 AAGGGCATCGCGCGCAACAGCAGATCCACAAAGCTGTGTGAGCAAGGGCATCCGCAAGTG 2349  
Db 704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723  
Qy 2350 CTGTTCTCTGGACCGCATCGAT 2370  
Db 724 LeuPheLeuAspGlyIleAsp 730  
RESULT 9  
GNLJSI  
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz  
N:Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly.  
C:Species: simian immunodeficiency virus SIVcpz  
A:Host: Pan troglodytes (chimpanzee)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C:Accession: S09984  
R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A:Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A:Reference number: S09983; MUID:90259077; PMID:2188136  
A:Accession: S09984  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1027 <HUE>  
A:Cross-references: UNIPROT:P17283; EMBL:X52154  
F:81-180/Product: retropepsin #status predicted <RTP>  
F:105/Active site: Asp (shared with dimeric partner) #status predicted  
C:Genetics:  
A:Gene: pol  
C:Superfamily: pol polypeptide  
C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleoside triphosphatase; peptidase  
F:81-180/Product: retropepsin #status predicted <RTP>  
F:105/Active site: Asp (shared with dimeric partner) #status predicted  
Alignment Scores:  
Pred. No.: 2,73e-142 Length: 1027  
Score: 3249.00 Matches: 599  
Percent Similarity: 89.46% Conservative: 63  
Best Local Similarity: 80.95% Mismatches: 75  
Query Match: 70.89% Indels: 3  
DB: 1 Gaps: 1  
US-09-610-313B-30 (1-2469) x GNLJSI (1-1027)  
Qy 159 CTGTGGAAGTGGCGCAAGGAGGCGCACCATGAGGACTGACCGAGCGCCAGGCCAA 218  
Db 7 LeuLeuAlaValTrpAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26  
Qy 219 CTCTTCCGCGAGACCTGCGCTTCCCGAGGCGAAGCGCGCGAGTTCCCGACGAGCA 278  
Db 27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGlu 46  
Qy 279 GAACCGCGCCCAACAGCCCGACCGCGAGCTGACGTCGCGCGC-----GACACCC 332  
Db 46 nAsnArgThrAsnGlyProThrAspArgGluLeuTrpValProGlyGlyArgGluPr 66  
Qy 333 CCGCAGCGAGCG 392  
Db 66 OGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86  
Qy 393 GCAGCGCCCTGTGTGAGCATCAAGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 452  
Db 86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106  
Qy 453 CG 512  
Db 106 rGlyAlaAspAspThrValIleGluArgIleGlnLeuGlnGlyLeuTrpLysProLysMe 126  
Qy 513 GATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 572  
Db 126 tIleGlyGlyIleGlyPheIleLysValLysGlnPheAsnValHisIleGluI 146



```
QY 481 ATGAGCCTGCGCCGACAGTGAAGCCCAAGATGATCGCGGCATCGCGGGCTTCATCAAG 540
Db 1 MetAenLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyPheIleLys 20
QY 541 GTGGCCAGTACGACCAAGATCCTCATCGAGATCTCGCGCAAGAGAGCCATCGGCACCGTG 600
Db 21 ValLysGlnTrpAspGlnIleAlaIleGluIleCysGlyHisLysAlaIleGlyThrVal 40
QY 601 CTGATCGGCCCCCGGTAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGC 660
Db 41 LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCys 60
QY 661 ACCGTGAACCTCCCATCGACCCCATCGAGACCTGCGCGTGAAGCTGAAGCCCGGCATG 720
Db 61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80
QY 721 GACGGCCCCAAGGTGAAGTGGCCCTGACCCAGGAGAGATCAAGGCCCTCACCGCC 780
Db 81 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuIleGlu 100
QY 781 ATCTCGAGGAGATGAGAGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTAC 840
Db 101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120
QY 841 AACACCCCGCTGTTCGCCATCAAGAGAGACAGACCAAGTGGCGCAAGCTGGTGGAC 900
Db 121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140
QY 901 TTCGGCAGCTGAACAGCCACCCAGGACTTCTCGGAGGTGACGCTGGGCATCCCCCAC 960
Db 141 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160
QY 961 CCGCGCCGCTGGAAGAGAGAGAGCGTACCGCTGCTGGAGCTGGCGCGACCGCTACTTC 1020
Db 161 ProGlyGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180
QY 1021 AGCGTGGCCCTGAGAGAGACTTCGCGAAGTACACCGCTTACCACCTCCAGCATCAAC 1080
Db 181 SerValProLeuAspLysAspPheArgLysTrpAlaPheThrIleProSerIleAsn 200
QY 1081 AACGAGACCCCGCATCCGCTACCTACAGTACAGTGTGCTGCCAGGGCTGGAGGGCAGC 1140
Db 201 AsnGluThrProGlyIleArgTrpGlnTrpAsnValLeuProGlnGlnTrpLysGlySer 220
QY 1141 CCAGACATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCGCAACCCC 1200
Db 221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240
QY 1201 GAGATCGTATCTACCAAGTACGACGACTGTACTGTGGCGAGCGACCTGGAGATCGGC 1260
Db 241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260
QY 1261 CAGCACCAGCCAGATCGAGGAGCTCGCAAGCAGCTGCGCTGGGGCTTCACCCACC 1320
Db 261 GlnHisArgThrLysIleGluGluLeuArgHisLeuLeuArgTrpGlyPheThrThr 280
QY 1321 CCGCAAGAGACCAAGAGAGCCCTTCTGTGGATGGCTTACGAGCTGCACCCC 1380
Db 281 ProAspLysLysHisGlnLysGluProProPheLeuTrpIleGlyTyrGluLeuHisPro 300
QY 1381 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACCGTGAACGAC 1440
Db 301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320
QY 1441 ATCCGAGAGCTGTGGCAGCTGMACTGGCCGACGATCTTACCCCGGCATCAAGTGTG 1500
Db 321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340
QY 1501 CGCCAGCTGTGCAAGCTGTGCGCGCGCCCAAGCCCTGACCCAGCATCGTGGCCCTGACC 1560
Db 341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360
QY 1561 GAGGAGCCGAGCTGGAGCTGGCGCGAGAACCGGAGATCTCGCGCGAGCCCGTGCACGGC 1620
```

```
Db 361 LysGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysThrProValHisGly 380
QY 1621 GTGTACTACCAACCCAGAGGACCTGGTGGCCGAGATCCAGAGAGCGAGGGCCACGACCAG 1680
Db 381 ValTyrTyrAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln 400
QY 1681 TGGACCTACCAAGTCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1740
Db 401 TrpThrTyrGlnIleTyrGlnProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420
QY 1741 ATCGCACCCCGCCACACCAAGCAGCTGAAGCAGCTGACCGGCGGTGCAGAAAGTCCGC 1800
Db 421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440
QY 1801 ATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAG 1860
Db 441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460
QY 1861 ACCTGGAGACCTGGTGGACCGACTACTTGGCAGGCCACCTGGATCCCGAGTGGAGTTC 1920
Db 461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480
QY 1921 GTGAACACCCCGCTGGTGAAGCTGTGTACCGCTGAGAGAGGAGCCCATCATCGGC 1980
Db 481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500
QY 1981 GCCGAGACCTTCTACGTGGACGGCGCCCAACCGCGAGACCAAGATCGGCAAGCCCGGC 2040
Db 501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520
QY 2041 TACGTGACCGACCGGGCCCGGACAGATCGTGAAGCTGACCGAGACCAACCAAGCAAG 2100
Db 521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys 540
QY 2101 ACCGAGCTGACGAGCATCTGAGCTGGCCCTGAGAGACGGCAGCGAGGTGAACATCGTG 2160
Db 541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560
QY 2161 ACCGACAGCCAGTACGCGCTGGGCATCATCAGCGCCCGACCGGCAAGAGCGAGCGAG 2220
Db 561 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 580
QY 2221 CTGGTGAACCAAGTCTGAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTG 2280
Db 581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600
QY 2281 CCCGCCACAGAGGCGCATCGCGGCAAGCAGCATCGACAAGCTGGTGAGCAAGGGCATC 2340
Db 601 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIle 620
QY 2341 CGCAAGGTGCTGTCTCTGGACGCGCATCGAT 2370
Db 621 ArgLysValLeuPheLeuAspGlyIleAsp 630
```

## RESULT 11

T01668

pol polyprotein - human immunodeficiency virus type 1

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C;Accession: T01668

R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.

Cell 46, 63-74, 1986

A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two iso

A;Reference number: Z14389, MUID:86245056; PMID:2424612

A;Accession: T01668

A;Status: translated from GB/EMBL/DDBU

A;Molecule type: mRNA

A;Residues: 1-902 &lt;Ali&gt;

A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230

C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	6.48e-137	Length:	902
Score:	3132.00	Matches:	578
Percent Similarity:	97.74%	Conservative:	28
Best Local Similarity:	93.23%	Mismatches:	14
Query Match:	68.34%	Indels:	0
DB:	2	Gaps:	0

US-09-610-313B-30 (1-2469) x T01668 (1-902)

QY	511	ATGATCGGCGCATCGCGGCTTCATCAAGGTGGCGCAGTACGACCGAGTCTCTGATCGAG	570
Db	1	MetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu	20
QY	571	ATCTGGCGCAAGAGCCATCGCACCGTCTGATCGGCCCCACCCCGTGAACATCATC	630
Db	21	IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle	40
QY	631	GGCGCAACATGTGACCCAGCTGGCTGCACCTGAACTTCCCATCGACCCCATCGAG	690
Db	41	GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	60
QY	691	ACCGTCCCTGAAGCTGAAGCCCGCATCGAGCGGCCCAAGGTGAAGCAAGTGGCCCTG	750
Db	61	ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTyrProLeu	80
QY	751	ACCGAGGAGAGATCAAGGCCCTGACGCCCATCTGCGAGGAGTGGAGAGGAGGCAAG	810
Db	81	ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys	100
QY	811	ATCACCAAGATCGGCCCGGAGAACCCCTACAAACACCCCGTTCGCGCATCAAGAGAAG	870
Db	101	IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys	120
QY	871	GACAGCACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGAACAGCGCACCCAGGAC	930
Db	121	AspSerThrLysTyrAspLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp	140
QY	931	TTCTGGAGGTGAGCTGGGCATCCCCACCCCGCGCGCTGAAGAGAGAGAGCGGTG	990
Db	141	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	160
QY	991	ACCGTGTGACGTGGCGCGCGCTACTTCAGCGTGGCCCTGGACGAGGACTTCCGCAAG	1050
Db	161	ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys	180
QY	1051	TACACCGCTTCACCATCCCGAGCATCAACAAAGAGACCCCGCGCATCCCTACAGTAC	1110
Db	181	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr	200
QY	1111	AACGTGTGCTGCCAGGGCTGMAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGCCAAG	1170
Db	201	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	220
QY	1171	ATCTGTGAGCCCTTCCCGCGCCGCAACCCCGAGATCGTGTATCTACAGTACATGGAGC	1230
Db	221	IleLeuGluProPheArgThrLysAsnProGluIleValIleIleTyrGlnTyrMetAsp	240
QY	1231	CTGTACTGGGCGCAGCACTGGAGATCGGCAGCACCGCGCCCAAGATCGAGAGCTGGCG	1290
Db	241	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	260
QY	1291	AAGCACTGTGCTGGGCTTCACCAACCCCGCACCAAGAGCAGCAGAGAGGAGCCCCC	1350
Db	261	GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	280
QY	1351	TTCTGTGGATGGGCTACGAGCTGCACCCGCAAGTGGACCGTGCAGCCCATCGAGCTG	1410
Db	281	PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu	300
QY	1411	CCGAGAGAGAGAGCTGGACCGGTGAACGACATCCAGAAAGCTGTGGGCAAGCTGGA	1470
Db	301	ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	320

QY	1471	GCCAGCCAGATCTACCCCGGCATCAAGGTGGCGCAGTGTGAAGCTGTGCGCGCGCC	1530
Db	321	AlaSerGlnIleIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla	340
QY	1531	AAGCCCTGACCGACATCGTGTGCCCTTGACCGAGAGGCGAGGTGGAGTGGCGCGAAC	1590
Db	341	LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn	360
QY	1591	CGCGAGATCTCGCGCGCGCGTGCACGGCTGTACTAGACCCCGACGAGGACCTGGTG	1650
Db	361	ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle	380
QY	1651	GCCGAGATCCAGAAAGCAGGCGCCACGACCACTGAGTGCACCTACAGATCTACAGAG	1710
Db	381	AlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGlnTyr	400
QY	1711	AGAACTGAAGACCGGCAAGTACGCAAGATCGCAGATCGCACCGCCACACCAACGAC	1770
Db	401	LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys	420
QY	1771	CAGCTCAGCGAGCGCTGCAGAGATCGCCATCGCAGAGATCGTGTGCGGCGGCAAGC	1830
Db	421	GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr	440
QY	1831	CCCAAGTTCGCGCTGCCATCCAGAAAGGAGACCTGGGAGACCTGTGTGGACCGACT	1890
Db	441	ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpGluTyrTrp	460
QY	1891	CAGGCCACTGTGGATCCCGAGTGGAGTTCGTGAAACACCCCGCTGTGTGAAGCTGTG	1950
Db	461	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp	480
QY	1951	TACACCTGGAGAGAGCGCCATCATCGCGCGCGAGACCTTCTACGTGAGCGCGCGCC	2010
Db	481	TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla	500
QY	2011	AACCGGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGACCGCGCGCGCAGAGATC	2070
Db	501	AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal	520
QY	2071	GTGAGCTGACCGAGACCCAAACCAAGAGACCGAGCTGCAGGCCATCCAGTGGCGCTG	2130
Db	521	ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu	540
QY	2131	CAGGACAGCGGCGCAGGCTGAACATCGTACCGCAGCAGCAGTACGCGCTGGCATCATC	2190
Db	541	GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle	560
QY	2191	CAGGCCAGCCCGCACAGAGCGAGCGAGCTGTGAAACCAAGATCATTCGAGCAGCTGATC	2250
Db	561	GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle	580
QY	2251	AGAAGAGAGAGTGTACTGTAGCTGGTGGTGGCGCCACCAAGGGCATCGCGCGCAAGC	2310
Db	581	GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu	600
QY	2311	CAGATCGACAGCTGTGTGACAGGCGCATCCGCAAGTGTGTCTTCTGGACGCGCATCGAT	2370
Db	601	GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp	620

RESULT 12

B47175  
reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1  
C;Species: human immunodeficiency virus type 1, HIV-1  
C;Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: B47175  
R;Mohri, H.; Singh, W.T.; Ho, D.D.  
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993  
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the  
A;Reference number: A47175; MUID:93126353; PMID:7678340  
A;Accession: B47175  
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual trar  
A;Molecule type: DNA

A;Residues: 1-559 <MOH>  
 A;Cross-references: UNIPROT:Q9PXX1  
 A;Note: sequence extracted from NCBI backbone (NCBIP:122099)  
 C;Superfamily: pol polyprotein

## Alignment Scores:

Pred. No.: 4,39e-122 Length: 559  
 Score: 2809.00 Matches: 518  
 Percent Similarity: 96.78% Conservative: 23  
 Best Local Similarity: 92.67% Mismatches: 18  
 Query Match: 61.29% Indels: 0  
 DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x B47175 (1-559)

```

Qy 673 CCCATCAGCCGCGGAGAGCCGCTCCGCTGAAGCTGAAGCCCGGCGATCGAGCGGCCCAAG 732
Db 1 ProleSerProfileGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20

Qy 733 GTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCGAGAG 792
Db 21 ValLysGlnTrpProLeuThrGluLysLysAlaLeuValGluLysCysThrGlu 40

Qy 793 ATGAGAGAGGAGGACAGATCACCAAGATCGGCCCGGAGACCCCTACACACCCCGTG 852
Db 41 MetGlnLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60

Qy 853 TTGCCCATCAAGAAAGAGGAGCAGCACCAAGTGGCGCAAGCTGTGGACTTCCGCGAGCTG 912
Db 61 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 80

Qy 913 AACAAAGCCGACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCACCCCGCGGCTG 972
Db 81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100

Qy 973 AAGNAGAGAGAGCGTGCAGCTGCTGAGCTGGCGGCGACCGCTACTTCAGGTGCCCTG 1032
Db 101 LysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120

Qy 1033 GACGAGACTTCCCAAGTACACCGCTTACCATCCCGAGCATCAACACGAGACCCCGC 1092
Db 121 AspGluAspPheArgLysThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140

Qy 1093 GGATCCGCTACCAAGTACACGCTGCTCCCGAGGGCTGGAAGGCGACCCCGAGCATTC 1152
Db 141 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160

Qy 1153 CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212
Db 161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIle 180

Qy 1213 TACCAGTACATGAGCAGACCTGTAGCTGGGAGCGACCTGGAGATCGGCGAGCAGCCGGCC 1272
Db 181 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluLysGlnHisArgThr 200

Qy 1273 AAGATCAGGAGCTGCGCAAGACCTGCTCGCTGGGGCTTACACCCCGCGACAAGAG 1332
Db 201 LysIleGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys 220

Qy 1333 CACCAGAGAGGCCCCCTTCTGTGTGATGGCTACGAGCTGCACCCCGGACAGTGAACC 1392
Db 221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240

Qy 1393 GTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAAAGCTG 1452
Db 241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260

Qy 1453 GTGGGCAAGCTGAACCTGGGCGAGCAGCATCTACCCCGGCGATCAAGGTGCGCGAGCTGTC 1512
Db 261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 280

Qy 1513 AAGCTGCTGCGCGCGCAAGGCCCTGACCGACATCGTGCCTGACCGAGAGCGCGAG 1572

```

```

Db 281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGlu 300
Qy 1573 CTGGAGCTGGCCGAGAACCCGCGAGATCTGCGCGAGCCCGTGACGGCGTGTACTACGAC 1632
Db 301 LeuGluLeuAlaGluAsnArgGluLeuLysGluProValHisGlyValTyrAsp 320
Qy 1633 CCCAGCAAGACCTGGTGGCGGAGATCCAGAGCAGGCGCCAGCAGCAGTGCATCACCAG 1692
Db 321 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGln 340
Qy 1693 ATCTTACCAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGGCACGCC 1752
Db 341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360
Qy 1753 CACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGAGAGATCGCCATCGAGAGCATC 1812
Db 361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380
Qy 1813 GTGATCTGGGCAAGACCCCAAGTTCGGCTCGCCATCCAGAGGAGACCTGGGAGACC 1872
Db 381 ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr 400
Qy 1873 TGGTGGACCGACTACTGCGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC 1932
Db 401 TrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420
Qy 1933 CCCCTGTGAAGCTGTGTACCTGACCTGAGAGAGAGCCCATCATCGGCGCGAGACCTTC 1992
Db 421 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 440
Qy 1993 TAGCTGACCGCGCGCCCAACCCCGGAGACCAAGATCCGCAAGGCGCGGTAGCTGACCGAC 2052
Db 441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460
Qy 2053 CGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGAGACCGAGCTGCAG 2112
Db 461 LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln 480
Qy 2113 GCATCCAGCTGGCCCTGCGAGGAGCGCGAGCGAGTGAACATCGTGACCGAGCCAG 2172
Db 481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500
Qy 2173 TAGCCCTGGGCATCATCCAGGCCCGCCGACCAAGAGCGAGCGAGCTGGTGAACCGAG 2232
Db 501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520
Qy 2233 ATCATCGAGCAGCTGTCAAGAGGAGAGTGTACCTGAGCTGGTGGCGCCGCCCAAG 2292
Db 521 IleIleGlnGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540
Qy 2293 GGATCCGCGCGCAACCGAGCAGATCGACAAGCTGTGAGCAAGGGCGCATCCGCAAGGTG 2349
Db 541 GlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559

```

## RESULT 13

A47175  
 reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1  
 C;Species: human immunodeficiency virus type 1, HIV-1  
 C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C;Accession: A47175  
 R;Mohr, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993  
 A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the  
 A;Reference number: A47175; PMID:93126353; PMID:7678340  
 A;Accession: A47175  
 A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A;Molecule type: nucleic acid  
 A;Residues: 1-559 <MOH>  
 A;Cross-references: UNIPROT:Q9PXX2  
 A;Note: sequence extracted from NCBI backbone (NCBIP:122100)  
 C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.: 3,66e-121 Length: 559  
 Score: 2789.00 Matches: 514  
 Percent Similarity: 96.60% Conservatives: 26  
 Best Local Similarity: 91.95% Mismatches: 19  
 Query Match: 60.86% Indels: 0  
 DB: 2 Gaps: 0

US-09-610-313B-30 (1-2469) x A47175 (1-559)

QY	673	CCCATGAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATGAGCGCCCAAG	732
Db	1	ProfileSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	20
QY	733	GTCAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGSCCTGACCGCCATCTCGGAGGAG	792
Db	21	ValLysGlnTrpProLeuThrGluLysIleValLysAlaLeuValGluIleCysThrGlu	40
QY	793	ATGGAGAAGGAGGCGAAGATCACCAAGATCGGCCCGGAGAACCCCTACAAACACCCCGTG	852
Db	41	MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal	60
QY	853	TTGCCATCAAGAAGAGGACAGACCAAGTGGCGCAAGCTGGTGACTTCCCGGAGCTG	912
Db	61	PheAlaIleLysLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu	80
QY	913	AACAAGCGCACCCAGGACTTCTGGAGGTGACGTGGGCATCCCGACCCCGCGCGCTG	972
Db	81	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	100
QY	973	AAGAAGAAGAGCGTGCCTGGACGTGGCGACGCGCTACTTACGCTGCGCCCTG	1032
Db	101	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	120
QY	1033	GAGAGGACTTCCGCAAGTACACGCGCTTCCACATCCCGAGCATCAACAACGAGACCCCG	1092
Db	121	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	140
QY	1093	GGATCCGCTACCAAGTACACGCTGCTCCCGAGGCTGGAAGGCGACGCCCGACCTTC	1152
Db	141	GlyLeuArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	160
QY	1153	CAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC	1212
Db	161	GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle	180
QY	1213	TACAGTACATGACGACCTGTAGCTGGGAGGACCTGAGATCGCGCAGCAGCCGCGCC	1272
Db	181	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr	200
QY	1273	AAGATCGAGGAGCTGCGCAAGCACCTGCTCGCTGGGGCTTCAACACCCCGCAAGAAG	1332
Db	201	LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	220
QY	1333	CACCAAGAAGGAGCCCTTCTGTGTGATGGGTACGAGCTGCACCCCGCAAGTGGACC	1392
Db	221	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	240
QY	1393	GTGAGCCCATCGAGCTGCGGAGAGGAGTGGACCTGACCGTGAACGACATCCAGAGCTG	1452
Db	241	ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu	260
QY	1453	GTGGGCAAGCTGAACCTGGCGCAGCATCTACCCCGGCATCAAGGTGCGCGCTGTC	1512
Db	261	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValLysGlnLeuCys	280
QY	1513	AAGTGTGCGCGCGCAAGGCCCTGACCGACATCGTCCCTGACCGGAGGCGCGAG	1572
Db	281	LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluGluAlaGlu	300
QY	1573	CTGAGCTGCGCGAGAACCGCGAGATCTCGCGGAGCCGTGCACCGCGTGTACTAGAC	1632
Db	301	LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp	320

QY	1633	CCCAAGCAAGACCTGGTGGCCGAGATCCAGAAGCAGGCGCACGACCGATGAGCTACCAG	1692
Db	321	ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGln	340
QY	1693	ATCTACAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCC	1752
Db	341	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla	360
QY	1753	CACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTCAGAAGATCGCCATGAGAGCATC	1812
Db	361	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrGluSerIle	380
QY	1813	GTGATCTGGCGGCAAGACCCCAAGTTCCGCTCCCATCCAGAAAGAGACCTGGAGACC	1872
Db	381	ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla	400
QY	1873	TGTTGACCGCATCTACCTGGCAGGCCACTGTGATCCCGAGTGGGAGTTCGTGAACACCCC	1932
Db	401	TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	420
QY	1933	CCCTCGTGAAGCTGTGTACCAAGTGGAGAGGCCCATCATCGCGCGGAGACCTTC	1992
Db	421	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe	440
QY	1993	TAGCTGGACGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTACACGAC	2052
Db	441	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	460
QY	2053	CGGGCGCGCAGAGATCGTGACCTGACCGAGACCAACCAACGAGACCGAGCTGCAG	2112
Db	461	LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln	480
QY	2113	GCCATCCAGCTGCGCTGCAGGACAGCGCAGGTGAACATCGTACCGACGACGACG	2172
Db	481	AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln	500
QY	2173	TACGCGCTGGGCATCTCCAGGCGCCCGCACAGAGCGAGCGAGCTGGTGAACACG	2232
Db	501	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln	520
QY	2233	ATCATCGACGCTGATCAAGAAGGAGAGGTGTACTGAGCTGGGTGCGCGCCCAAG	2292
Db	521	IleIleGluGluLeuIleLysLysGlyValTyrLeuAlaTrpValProAlaHisLys	540
QY	2293	GGCATCGGCGGCAACGAGCAGATCGCAAGCTGCTGAGCAAGGCGCATCCCGAAGGTG	2349
Db	541	GlyIleGlyLysAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal	559

RESULT 14

S46347  
 pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)  
 C;Species: simian immunodeficiency virus SIVagm  
 A;Variety: isolate SAB-1  
 C;Date: 25-Dec-1994 #sequence\_revision 14-Feb-1997 #text\_change 26-Aug-1999  
 C;Accession: S46347  
 R;Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.; et al.  
 EMBO J. 13, 2935-2947, 1994  
 A;Title: Mosaic genome structure of simian immunodeficiency virus from West African green  
 A;Reference number: S46335; MUID:94298785; PMID:8026477  
 A;Accession: S46347  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-1039 <JIN>  
 A;Cross-references: EMBL:U04005; NID:9466229; PIDN:AAA21505.1; PID:9466231  
 A;Experimental source: isolate SAB-1; sabaeus monkey  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
 C;Genetics:  
 A;Gene: pol  
 C;Superfamily: pol polyprotein  
 Alignment Scores: 1.13e-102 Length: 1039  
 Pred. No.:

Score: 2388.00 Matches: 444  
Percent Similarity: 74.60% Conservative: 120  
Best Local Similarity: 58.73% Mismatches: 150  
Query Match: 52.11% Indels: 42  
DB: 2 Gaps: 6

US-09-610-313B-30 (1-2469) x S46347 (1-1039)

QY 220 TTCTTCCGCGAGGACCTGGCTTCCCGAGGCG-----AAGCGCGCGAGTTCCTCCCGAGC 273  
Db 1 PhePheArg-----ValTrpProLeuGlnArgGluThrGlnGluPheProSer 17  
QY 274 GAGCAGAACCGCGCCAAAC---AGCCCCACCGAGCGCGAGCTGCAGTGGCGGC----- 324  
Db 18 AspleuHisGlnThrAsnSerSerProAsnGlyThrGlyLeuGlnGlnAlaGlyGlyLys 37  
QY 325 -----GACAACCCCGCGAGGCGC----- 345  
Db 38 LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgSerSerAsnSerPro 57  
QY 346 -----GGCGCGGAGCGC----- 357  
Db 58 VallysAlaValCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla 77  
QY 358 -----CAGGCGACCTGAACCTTCCCGAGATCACCCTGTGGCAGCGCCCC 402  
Db 78 ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgArgPro 97  
QY 403 CTGTGTGACATCAAGTGGCGGCGCGAGATCAAGGAGGCGCTGTGGACACCGCGCGCGAC 462  
Db 98 MetIysThrValTrpIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp 117  
QY 463 GACACCGTGTGGAGAGATGAGCTGCCCGGCAAGTGGAGCCCAAGATGATCGCGCGC 522  
Db 118 AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly 137  
QY 523 ATCGCGGCTTCATCAAGTGGCGCGAGTACGACGAGTCTGTGATCGAGATCTGGCGAAG 582  
Db 138 IleGlyGlyCysIleAsnValLysAlaTrpHisAsnGlnGluValLysIleGluAspLys 157  
QY 583 AAGGCGATCGGCGCGCTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 642  
Db 158 ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal 177  
QY 643 CTGACCCAGCTGGGTGACCTGACCTTCCCGATCCCGCGCGCGCGCGCGCGCGCGCGCG 702  
Db 178 LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal 197  
QY 703 AAGTGAAGCG 762  
Db 198 HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys 217  
QY 763 ATCAAGCG 822  
Db 218 IleGluAlaLeuLysAlaIleCysGluAspLeuGluLysGlnGlyHisLeuGluArgIle 237  
QY 823 GGCGCGCGAGAACCTTACCAACACCGCGTGTTCGCGCATCAAGAAGAGGAGGAGGAGGAG 882  
Db 238 GlyProGluAsnProTrpAsnThrProValPheAlaIleArgLysLysAspLysThrGln 257  
QY 883 TGGCGCAAGCTGTGGATCTCGGAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 942  
Db 258 TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal 277  
QY 943 CAGTGGCGATCCCG 1002  
Db 278 GlnLeuGlyIleProHisProAlaGlyLeuGlnArgGluGlnIleThrValLeuAsp 297  
QY 1003 GTGGCGCGAGCGCTTACTTCAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1062  
Db 298 IleGlyAspAlaTrpPheSerCysProLeuAspProAspPheGlnLysTrpAlaPhe 317  
QY 1063 ACCATCCCGCGATCAACACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1122

Db 318 ThrIleProSerValAsnAsnArgGluProGlyIleArgTrpGlnTrpLysValLeuPro 337  
QY 1123 CAGGCTCGAAGGCGAGCG 1182  
Db 338 GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu 357  
QY 1183 TTCCG 1242  
Db 358 PheArgGlnLysAsnProAspValAspIleTrpGlnTrpMetAspAspMetLeuIleAla 377  
QY 1243 AGCGACCTGGAGATCG 1302  
Db 378 SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTrpLeuGlu 397  
QY 1303 CGCTGGCGCTTCCACCG 1362  
Db 398 ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProProTrpLeuTrpMet 417  
QY 1363 GGCTACGAGTGCACCG 1422  
Db 418 GlyTrpGluLeuTrpLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu 437  
QY 1423 AGCTGCG 1482  
Db 438 GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 457  
QY 1483 TACCG 1542  
Db 458 TyrThrGlyIleLysThrLysHisLysCysArgLeuIleArgGlyAlaArgProLeuThr 477  
QY 1543 GACATCGTGGCG 1602  
Db 478 GluIleValGlnTrpThrGluAlaGluLeuGluLeuGluLysArgGlnIleLeu 497  
QY 1603 CGCGAGCGCGTGCAGCGCGTGTACTAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1662  
Db 498 ArgGlnLysGlnGlnGlyGlnTrpTyrAspProAlaLeuProLeuArgAlaLysValLeu 517  
QY 1663 AAGCAGCG 1722  
Db 518 LysLeuGlyAspGlyGlnTrpGlyTrpGlnIleTrpGlnProGluAsnLysIleLeuLys 537  
QY 1723 ACCGCGAGTACCG 1782  
Db 538 ValGlyLysTrpAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly 557  
QY 1783 GCGTGCAGAGATCG 1842  
Db 558 LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu 577  
QY 1843 CTGCGCGCATCCAGAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1902  
Db 578 LeuProValGluArgGluLeuTrpGluGlnTrpTrpSerAspTrpTrpGlnValThrTrp 597  
QY 1903 ATCCCGCGAGTGGAGTTCGTGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1962  
Db 598 IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTrpLysLeuVal 617  
QY 1963 AAGGAGCG 2022  
Db 618 LysAspProIleProGlyGluAlaValTrpValAspGlyAlaAlaAsnArgAsnSer 637  
QY 2023 AAGTTCG 2082  
Db 638 LysGluGlyLysAlaGlyTrpLeuThrAspArgGlyAspGlnLysValValAlaLeuGlu 657  
QY 2083 GAGACCAACCAAGAGACCG 2142  
Db 658 AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly 677  
QY 2143 AGCGAGGTGAACATCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2202



Db 678 SerLysValAsnIleIleThrAspSerGlnTyRAlaMetGlyIleIleAlaGlyGluPro 697  
QY 2203 GACAAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAGGAGAAG 2262  
Db 698 ThrGluSerAspAsnAsnIleValGlnGlnIleIleGluLeuIleIleLysGluAla 717  
QY 2263 GTGTACCTGAGCTGGTGGCCGCCCAACAAGGGCATCGCGGCCAACGAGCAGATCGACAAG 2322  
Db 718 ValTyRileAlaTrpValProAlaHisLysGlyValGlyGlyAsnGluGluIleAspLys 737  
QY 2323 CTGGTAGCAAGGCGATCCGCAAGGTGCTTCTGCGAGCGCATCGAT 2370  
Db 738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753  
RESULT 15  
S53092  
pol polyprotein - human immunodeficiency virus type 2  
C;Species: human immunodeficiency virus type 2, HIV-2  
C;Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C;Accession: S53092  
R;Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.  
Submitted to the EMBL Data Library, March 1995  
A;Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu  
A;Reference number: S53091  
A;Accession: S53092  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1055 <REC>  
A;Cross-references: UNIPROT:O73194; EMBL:Z48731  
C;Superfamily: pol polyprotein  
C;Keywords: polyprotein  
Alignment Scores:  
Pred. No.: 4,8e-96 Length: 1055  
Score: 2244.00 Matches: 421  
Percent Similarity: 71.11% Conservative: 128  
Best Local Similarity: 54.53% Mismatches: 184  
Query Match: 48.96% Indels: 39  
DB: 2 Gaps: 9  
US-09-610-313B-30 (1-2469) x S53092 (1-1055)

QY 156 GGGCTGTGGAAGTGGCGGCAAGGAGGCCACACAGATGAAGGACTGCAC----CGAGCGCCA 212  
Db 3 GlyLeuLeuGluMetTrpGln-----AspArgThrTyRHisGlyLysValPro 18  
QY 213 GGCCAACTT-----CTCCG 227  
Db 19 ArgGlnThrGlyGlyPhePheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38  
QY 228 CGAGGACCTGGCTTCCCCAGGGCAAGGCCGCGAGTTCGCCAGCGAGCAGAA----- 281  
Db 39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerGly 58  
QY 282 CCGCGCCAACAGCCCAACAGCCGCGCA-----GCTGCAGGTGGCGCGCACCAACCCCGC 335  
Db 59 ProThrGlyGluIleTyRAlaAlaArgGluLysAlaGluAlaGluArgGluThrIle 78  
QY 336 CAGCAGCGCGCGC-----CGAGCGCCA-----GGGCACCCCTGAACTT----- 374  
Db 79 GlnArgGlyAspArgGlyLeuAlaAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98  
QY 375 -----CCCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAA 416  
Db 99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValThrAlaHi 118  
QY 417 GGTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGACACACCGTGTCTGGA 476  
Db 118 sIleGluGlyGlnProValGluValLeuLeuAspThrGlyAlaAspSerIleValAl 138  
QY 477 GAGATGAGCTGCGCGCGCAAGTGAAGCCACAGATATCGCGGCGATCGCGCGCTTCAT 536  
Db 138 agIyIleGluLeuGlySerAsnTyRProLysIleValGlyIleGlyIlePheIl 158

QY 537 CAAGGTGCGCCAGTACGACACAGATCCTGATCGAGATCTCGGCAAGAAGCCATCGGCAC 596  
Db 158 eAsnThrLysGluTyRLeuLysAsnValGluIleGluValLeuGlyLysArgValargAlaTh 178  
QY 597 CGTGTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATGTCGACCCAGCTGGG 656  
Db 178 rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuG 198  
QY 657 CTGCACCCCTGAACTTCCCATCAGCCCCATCAGACCGTGCCTGAAGCTGAAGCCCG 716  
Db 198 yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProG 218  
QY 717 CATGACGCGCCCAAGCTGAAGCAGTGGCCCTCGACCGGAGAGAACATCAAGGCCCTGAC 776  
Db 218 yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeu 238  
QY 777 CGCATCTCGAGAGATGAGAGGAGGAGGCAAGATACCAAGATCGGCCCGGAGAACCC 836  
Db 238 sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluGluAlaProProThrAsnPr 258  
QY 837 CTACAAACCCCGCTGCTTCCCATCAAGAGAGAGGACAGCACCAAGTGGCGCAAGCTGT 896  
Db 258 ofrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeuIl 278  
QY 897 GGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGCGCATCCC 956  
Db 278 eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298  
QY 957 CAACCCCGCGCGCTGAAGAAGAGAGCGTGACCGTCTCGAGCTGGCGGACGCGCTA 1016  
Db 298 oHisProAlaGlyLeuAlaLysLysArgArgIleThrValLeuAspValGlyAspAlaTy 318  
QY 1017 CTTACGGTGCCTCGAGGAGACTTCCGAGTACACCGCTTCCACCATCCCCAGCAT 1076  
Db 318 rPheSerIleProLeuHisGluAspPheArgGlnTyRThrAlaPheThrLeuProSerVa 338  
QY 1077 CAACAACGAGACCCCGCGCATCCGCTACCAAGTCAACAACGCTGCTCCCGGAGGCTGGAAGG 1136  
Db 338 lAsnAsnAlaGluProGlyLysArgTyRileTyLysValLeuProGlnGlyTrpLysGl 358  
QY 1137 CAGCCCGAGATCTTCAGAGCAGCATGACCAAGATCTTGGAGCCCTTCCGCGCGCCGCAA 1196  
Db 358 ySerProAlaIlePheGlnTyRThrMetArgGlnValLeuGluProPheArgLysAlaAs 378  
QY 1197 CCGGAGATCGTATCATCAGTACATGAGCAGCTGTACGTCGGGAGCGACCTGGAGAT 1256  
Db 378 nGlnAspValIleIleIleGlnTyRMetAspAspIleLeuIleAlaSerAspArgThrAs 398  
QY 1257 CGCCACGACCCGCGCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGTGGGCTTCAC 1316  
Db 398 pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuLeuAsnSerLeuGlyPheSe 418  
QY 1317 CACCCCGCAGACAAGACACAGAGAGCCCGCTTCTGATGGCTGATGGCTACGAGCTGCA 1376  
Db 418 rThrProAspGluLysPheGlnLysAspProProTyArgTrpMetGlyTyRLeuLeuTr 438  
QY 1377 CCGGACAGTGGACCGCTGCGAGCCCATCGAGCTCCCGAGAGAGAGAGCTGGACCGTAA 1436  
Db 438 pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysIleValTrpThrValAs 458  
QY 1437 CGACATCCAGAAGCTGTGGCAAGCTGAACCTGGCGCAGCCAGATCTTACCCCGGCTCAA 1496  
Db 458 nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyRProGlyIleLy 478  
QY 1497 GTGCGCGCAGCTGTGAAGCTGTGCGCGCGCGCAAGGCCCTGACCGACATCGTGGCCCT 1556  
Db 478 sThrLysHisLeuCysArgLeuIleArgGlyLysMetThrLeuThrGluIleGlnTr 498  
QY 1557 GACCGAGGAGCGCGCTGAGCTGCCGAGAACCGCGAGATCCTGCGCGAGCCCGGTGCA 1616  
Db 498 pThrGluLeuAlaGluAlaGluLeuGluAsnArgValIleLeuSerGlnGlnGln 518



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 11:33:41 ; Search time 168.242 Seconds  
(without alignments)  
11351.628 Million cell updates/sec

Title: US-09-610-313B-30  
Perfect score: 4583  
Sequence: 1 gtgacccaccatggcgca.....gggctagaccggtgaattc 2469

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPRO.spool/US09610313/runat\_31052005\_15136\_15139/app.query.fasta\_1.7893  
-DB-A Geneseq 16Dec04 -OPMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09610313 @CGN 1 1 557 @runat\_31052005\_15136\_15139 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPOP=6  
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 16Dec04: \*  
1: Geneseqp1980s: \*  
2: Geneseqp1990s: \*  
3: Geneseqp2000s: \*  
4: Geneseqp2001s: \*  
5: Geneseqp2002s: \*  
6: Geneseqp2003as: \*  
7: Geneseqp2003bs: \*  
8: Geneseqp2004s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3804.5	83.0	1435	ADS19486	ADS19486 Modified
2	3782.5	82.5	3183	ADP84803	ADP84803 HIV-1 hyb
3	3741.5	81.6	1457	ADN36406	ADN36406 Human pro
4	3658	79.8	1003	AAB69289	AAB69289 HIV-1 non
5	3649	79.6	999	3 AAB69286	AAB69286 HIV-1 non
6	3644	79.5	1005	3 AAB69287	AAB69287 HIV-1 non
7	3608	78.7	998	6 AAE37601	AAE37601 HIV-1 sub
8	3601	78.6	3025	4 AAB86169	AAB86169 HIV-1 sub
9	3555.5	77.6	1000	3 AAB69282	AAB69282 HIV-1 non
10	3537.5	77.2	1002	2 AAW72993	AAW72993 HIV isola

11	3537	77.2	854	5	AAM48949	Aam48949 HIV-1 sub
12	3535	77.1	1003	6	AAO30963	Aao30963 HIV pol p
13	3535	77.1	1003	8	ADN36413	Adn36413 HIV prote
14	3532.5	77.1	1002	1	APB1861	Apb1861 Sequence
15	3532	77.1	1003	6	ABR55489	Abt55489 Amino aci
16	3530	77.0	1003	1	APB60420	Apb60420 Sequence
17	3530	77.0	1003	1	APB70861	Apb70861 Sequence
18	3524	76.9	1003	3	AAI70602	Aai70602 Codon opt
19	3524	76.9	1003	3	AAI70601	Aai70601 Corrected
20	3522	76.8	1003	5	AAO19387	Aao19387 Lymphaden
21	3516.5	76.7	1005	1	APB08060	Apb08060 HIV-1 pol
22	3513	76.7	1015	1	APB60347	Apb60347 HTLV-III
23	3513	76.7	1015	2	AAK43867	Aak43867 HTLV-III
24	3513	76.7	1015	4	AAH85993	Aah85993 Amino aci
25	3510	76.6	1003	3	AAH10047	Aah10047 HIV-1 pol
26	3510	76.6	1003	3	AAI70600	Aai70600 Wild type
27	3509	76.6	1016	2	AAO8062	Aao8062 ACNPV-HIV
28	3506	76.5	1015	2	AAK43875	Aak43875 HTLV-III
29	3502	76.4	1016	2	AAO8063	Aao8063 HIV-1 pol
30	3499	76.3	2033	2	AAO8056	Aao8056 HIV-1 pol
31	3495	76.3	1016	2	AAO8054	Aao8054 HIV-1 pol
32	3493	76.2	1003	3	AAB69284	Aab69284 HIV-1 non
33	3492	76.2	2033	2	AAO8055	Aao8055 HIV-1 pol
34	3491	76.2	1003	8	ADP20076	Adp20076 Human imm
35	3487	76.1	1003	1	AAH61508	Aah61508 Sequence
36	3487	76.1	1003	2	AAO29705	Aao29705 pol gene
37	3487	76.1	1003	3	AAI77301	Aai77301 HIV-1 (AT
38	3487	76.1	1003	5	AAE35789	Aae35789 ARV-2 (9B
39	3487	76.1	1003	6	ABU63185	Abu63185 Pol prote
40	3487	76.1	1491	1	APB1048	Apb1048 Transcrip
41	3484	76.0	1003	3	AAH69279	Aah69279 HIV-1 non
42	3482	76.0	739	5	AAU11874	Aau11874 HIV pol p
43	3482	76.0	1003	6	ABU57552	Abu57552 AIDS asso
44	3480	75.9	1012	6	ABU63325	Abu63325 Human lym
45	3477.5	75.9	1003	2	AAO8059	Aao8059 HIV-1 pol

## ALIGNMENTS

RESULT 1  
ADS19486  
ID ADS19486 standard; protein; 1435 AA.

AC ADS19486;  
XX  
DT 18-NOV-2004 (first entry)  
XX  
DE Modified HIV-1 GagPol protein.  
XX  
KW Gene delivery; epithelial cell; respiratory system; mammal;  
KW lentiviral packaging system; expression vector; gagpol; gene expression;  
KW lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;  
KW cystic fibrosis; CF; deltaPD; transmembrane potential difference;  
KW gene therapy; HIV-1; mutant; mutein.  
XX  
OS Human immunodeficiency virus 1.  
OS Synthetic.  
XX  
PN US2004037780-A1.  
XX  
PD 26-FEB-2004.  
XX  
PF 23-AUG-2002; 2002US-00226638.  
XX  
PR 19-NOV-2001; 2001AU-00008942.  
XX  
PA (PARS/) PARSONS D.  
PA (ANSO/) ANSON D.  
PA (LIMB/) LIMBERIS M.  
PA (FULL/) FULLER M.  
XX  
PI Parsons D, Anson D, Limberis M, Fuller M;  
XX





Qy	13	ATGCGCGAGGCCATGAGCCAGGCCACCAGC---GCCAACATCTCTGATCGAGCGAGCAAC	69
Db	363	LeuAlaGluAlaMetSerGlnValThrAenProAlaThrIleMetIleGlnLysGlyAsn	382
Qy	70	TTCAAGGGCCCCAAGGCGCATCATCAAGTCTCTCACTCGCGGCAAGAGGGCCACATCGCC	129
Db	383	PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla	402
Qy	130	CGCAACTGCGCGCCCGCCCGCAAGAGGGCTCTCGAAGTGGCGGCAAGAGGGCCACCAG	189
Db	403	LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln	422
Qy	190	ATGAAGGACTGCACCGCGCCAGGCCCAAC-----	219
Db	423	MetLysAspCysThrGluArgGlnAlaAenPheLeuGlyLysIleTrpProSerHisLys	442
Qy	220	-----TTCTTCGGCAGGAC	234
Db	443	GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp	462
Qy	235	CTGGCTTCCCCAGGGCAAGCCCGGAGTTCCCCAGCGAGCAACCGCGCCACAGC	294
Db	463	LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGlnThrArgAlaAsnSer	482
Qy	294	-----	294
Db	483	ProProGluSerPheArgPheGlyGluGluThrThrThrProSerGlnLysGlnGlu	502
Qy	295	-----CCACAGCGCGAGCTGCAGTG-----	318
Db	503	ProIleAspLysGluLeuTyProLeuAlaProThrArgArgGluLeuGlnValTrpGly	522
Qy	319	CGGGCGCAACCCCGCAGCGAGGGCGGGCGGAGCGCCAGGGCACCTGAAC-----	372
Db	523	ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer	542
Qy	373	-----TTCCCCCAGATCACCTGTGG	393
Db	543	SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp	562
Qy	394	CAGCGCCCTGTGTGATCAAGTGGCGGCCAGATCAAGGAGGCCCTCTCGACACC	453
Db	563	GlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThr	582
Qy	454	GGCGCGACACCGTGTGTGGAGGATGAGCTGCCCGGCAAGTGGAGGCCAAGATG	513
Db	583	GlyAlaAspAspThrValLeuGluGluMetAsnLeuLeuProGlyArgTrpLysProLysMet	602
Qy	514	ATCGGGGCATCGCGCTTTCATCAAGTGGCGCAGTAGCAGCCAGATCTGTATCGAGATC	573
Db	603	IleGlyGlyGlyGlyPheIleLysValArgGlnTyArgPheGlnIleLeuIleGluIle	622
Qy	574	TGGCGCAAGAGCCATCGGCACCGTGTGATCGCGCCCGCCACCCCGTGAACATCATCGGC	633
Db	623	CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly	642
Qy	634	CGCAACATGTGACCCAGCTGGGCTGCACCTTGAATCTTCCCATCAGCCCGCATCGAGACC	693
Db	643	ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr	662
Qy	694	GTGCGGTGAAGCTGAAGCCCGCATGGAGCGGCCCGCCAGGTGAAGCAGTGGCCCTGACC	753
Db	663	ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr	682
Qy	754	GAGGAGAAGATCAAGGCCCTTGACCGCATCTGCGAGGAGATGGAGAAGGAGGCAAGATC	813
Db	683	GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle	702
Qy	814	ACCAAGATCGGCGCCGAGAACCCCTTACAACACCCCGTGTTCGCCATCAAGAAGAGGAC	873
Db	703	SerLysIleGlyProGluAsnProTyArgThrProValPheAlaIleLysLysAsp	722



Db 1083 GlnLeuGluLysGluProIlelleGlyAlaGluThrPheTyrValAspGlyAlaAlaAsn 1102  
Qy 2014 CGCGAGCAAGATGGCAAGCCGGCTACCTGACCGACCGCGCGCGCAAGATCGTG 2073  
Db 1103 ArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValVal 1122  
Qy 2074 AGCTGTACCGAGACCAACAGACAGACAGACAGACAGACAGACAGACAGACAG 2133  
Db 1123 ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142  
Qy 2134 GACAGCGCAGCGAGGTGAACATCGTGACACAGACAGACAGACAGACAGACAG 2193  
Db 1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGln 1162  
Qy 2194 GCCAGCGCCGCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAG 2253  
Db 1163 AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuLys 1182  
Qy 2254 AAGGAGAAGGTGTACTGAGCTGGGTGCGCCGCCACACAGGCGCATCGCGGCAAGAGCAG 2313  
Db 1183 LysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGln 1202  
Qy 2314 ATCCAGCAAGCTGCTGAGCAAGCGCATCGCAAGGTGCTGTTCTTGGACGGCATCGAT 2370  
Db 1203 ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221  
  
RESULT 3  
ADN36406  
ID ADN36406 standard; protein; 1457 AA.  
AC ADN36406;  
XX  
XX 15-JUL-2004 (first entry)  
XX  
XX Human protein for anti-HIV vaccine.  
XX  
XX anti-HIV; vaccine; HIV; promoter; viral particle; immunization.  
XX  
XX Homo sapiens.  
XX  
XX WO2004035006-A2.  
XX  
XX 29-APR-2004.  
XX  
XX 17-OCT-2003; 2003WO-US033112.  
XX  
XX 18-OCT-2002; 2002US-0419465P.  
XX  
XX (AARO-) AARON DIAMOND AIDS RES CENT.  
XX  
XX Huang Y, Ho DD, Chen Z;  
XX  
XX WPI; 2004-348328/32.  
XX  
XX N-PSDB; ADN36405.  
XX  
XX Nucleic acid vector comprising at least one HIV sequence operably linked  
XX to a promoter and encoding a protein that does not assemble into viral  
XX particles, useful in immunizing a subject against HIV infection.  
XX  
XX Disclosure; SEQ ID NO 20; 166pp; English.  
XX  
XX The invention relates to a nucleic acid vector comprising at least one  
XX HIV sequence operably linked to a promoter and encoding a protein that  
XX does not assemble into viral particles. The nucleic acid vector is useful  
XX in immunizing a subject against HIV infection. This sequence corresponds  
XX to a peptide used in the invention.  
XX  
XX Sequence 1457 AA;  
  
Alignment Scores: 4.08e-211 Length: 1457  
Pred. No.: 3741.50 Matches: 709  
Score:

Percent Similarity: 92.24% Conservative: 28  
Best Local Similarity: 88.74% Mismatches: 44  
Query Match: 81.64% Indels: 19  
DB: 8 Gaps: 3  
US-09-610-313B-30 (1-2469) x ADN36406 (1-1457)  
Qy /13 ATGCGCAGCCATGAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAG 72  
Db 383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401  
Qy 73 AAGGGCCCCAAGCGCATCATCAAGTGTTCACCTGCGCGCAAGAGAGGCGCAGCATCGCCGCG 132  
Db 402 LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421  
Qy 133 AACTGCGCGCCCGCCCAAGAGAGGCTGTGGAGTGCAGCAAGAGAGGCGCAGCATG 192  
Db 422 AsnCysArgAlaProAspLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 441  
Qy 193 AAGCACTGCACCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAG 252  
Db 442 LysAspCysThrGluArgGlnAlaAsnPheLeu---GlyLysIleTrpProProHisLysG 461  
Qy 253 AAGCGCGCGAGTTCGCCAGCGAGCAGAAACCGCGCCCAACAGCGCAGCGCAGCGAGCTG 312  
Db 461 YArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPh 481  
Qy 313 CAGGTGCGCGCGCAACACCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372  
Db 481 eGlyPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501  
Qy 373 TTC-----CCCCAGATCACCCTGTGGCA 395  
Db 501 rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpGl 521  
Qy 396 GCGCGCCCTGCTGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455  
Db 521 nArgProLeuValSerIleArgValGlyGlyGlnIleLysGlnAlaLeuLeu----- 538  
Qy 456 CGCGCAGCACACCGTGTGTGGAGGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAGATGAT 515  
Db 539 ----AspAspThrValLeuGluValAsnLeuProGlyLysTrpLysProLysMetIl 557  
Qy 516 CGCGCGCATCGCGCGCTTCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 575  
Db 557 eGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleProIleGluIleCy 577  
Qy 576 CGGCAAGAGCGCCATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 635  
Db 577 sGlyLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597  
Qy 636 CAACATGCTGACCGCGCTGCGCTGCACCTGAACTTCCCATCAGCGCCCATCGAGACCGGT 695  
Db 597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrIl 617  
Qy 696 GCCCGTCAAGCTGAAGCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755  
Db 617 eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGl 637  
Qy 756 GGAGAAATCAAGCCCTGACCGCGCATCTGGAGGAGATGAGAGAGGAGGAGGAGGAGGAG 815  
Db 637 uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657  
Qy 816 CAAGATCGCGCGCGCGAGAACCCCTACAAACACCGCGCGTTCGCCCATCAAGAGAGAGGAG 875  
Db 657 rLysIleGlyProGlnAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677  
Qy 876 CACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGCAACAGCGCAGCGCGCGCGCGCT 935  
Db 677 rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697  
Qy 936 GGAGTGCAGTGGCGCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995

697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVa 717  
996 GCTGGAGCTGGCGGACCCCTACTTCAGCGTGCCTCGAGAGAGACTTCGCGAAGTACAC 1055  
717 lLeuAspValGlyAspAlaTyrPheSerValProLeuTyrGluAspPheArgLysTyrTh 737  
1056 CGCCTTTACCATCCCCAGCATCAACACAGAGACCCCGGCATCCGCTACCAGTACAACTG 1115  
737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVa 757  
1116 GCTGCCCCAGGCTGGAGGGAGCCCCAGCATCTTCCAGAGCAGATGACCAAGATCCT 1175  
757 lLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777  
1176 GGAGCCCTCCCGCGCCGCAACCCCGAGATCGTGATCTACCTAGTACAGACGACCTGTGA 1235  
777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyrGlyAspAspLeuTyr 797  
1236 CGTGGGAGGAGCTGAGATCGGCCAGCACCCGCGCCCAAGATCGAGGAGCTGGCGCAAGCA 1295  
797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHI 817  
1296 CTTGCTGGCTGGGGCTTACACCCCGCGAGAGACCAAGAGAGCCGCCCTTCCT 1355  
817 sLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLe 837  
1356 GTGATGGGTACGAGTGCACCCGACAGTGCAGCGTGCAGCCATCGAGCTGCCCGA 1415  
837 utrMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGI 857  
1416 GAAGGAGAGCTGGACCGTGAACACATCCAGAGCTGGTGGGCAAGTGAACCTGGGCCAG 1475  
857 uLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877  
1476 CCGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTCTGCGCGCGCCAGGC 1535  
877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897  
1536 CTTGACGAGCATGTCGCCCTGACCGAGAGGCGGAGCTGGAGCTGGCGAGAACCGCA 1595  
897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGI 917  
1596 GATCTCGCGGAGCCGTCACCGCGCTGTACTACGACCCCGACAGGACCTGTGTCGCCGA 1655  
917 uIleLeuLysGluProValHisGlyAlaTyrTyrAspProSerLysAspLeuIleAlaGI 937  
1656 GATCCAGAACGAGGCCACACAGATGGACCTTACAGATCTACAGAGCCCTTCAAGAA 1715  
937 uIleGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAs 957  
1716 CTTGAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGAGCTGAAGCAGCT 1775  
957 nLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977  
1776 GACGAGGCGCTGCAGAAATCCATCGCGCCGAGATCGTATCTGGGCGAAGACCCGCCAA 1835  
977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTrpGlyLysIleProLy 997  
1836 GTTCCGCTCCCATCCAGAGAGAGACTGGGAGACTGTGTGACCGGACTTCTGGCAGGC 1895  
997 sPheArgLeuProIleProLysGluThrTrpGluThrArgTrpThrAlaTyrTrpGlnAl 1017  
1896 CACTCGATCCCGAGTGGAGTTCTGTGACACCCCCCTCGTGAAGCTGTGTACCA 1955  
1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGI 1037  
1956 GCTGGAGAGAGGCCCATCATCGCGCCGAGACTTCTAGCTGGAGCGCGCCGCCAACCG 2015  
1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyrValAspGlyAlaAlaAsnAr 1057  
2016 CGAGACCAAGATCGGCAAGCCCGCTACGTGACCGACCGCGCGCGGCGGAGAGTCTGTGAG 2075  
1057 gGluThrLysMetGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSe 1077

2076 CTTGACCGAGACCAACCAAGAGACCGAGCTGAGGCCATCCAGCTGGCCCTGCAGGA 2135  
1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097  
2136 CAGCGCGCAGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGATCATCCAGGC 2195  
1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAl 1117  
2196 CCAGCCCGACAGAGCAGAGCGAGCTGCTGAACACAGATCATCGACGACGTGATCAAGAA 2255  
1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuLysLy 1137  
2256 GGAGAAGGTGTACTCGAGCTGGGTGCCGCCACAGGGCATCGCGGCAACCGAGCAGAT 2315  
1137 sGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVa 1157  
2316 CGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCCTGGACGGCATCGAT 2370  
1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175

RESULT 4  
AAB69289  
ID AAB69289 standard; protein; 1003 AA.  
XX AAB69289;  
XX  
DT 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX  
DE HIV-1 non-subtype B clone 94IN476-104 pol protein.  
XX  
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX  
OS Human immunodeficiency virus 1.  
XX  
PN WO200026416-A1.  
XX  
PD 11-MAY-2000.  
XX  
PF 25-OCT-1999; 99WO-US024837.  
XX  
PR 02-NOV-1998; 98US-00184418.  
XX  
XX (UABR-) UAB RES FOUND.  
XX  
XX Hahn BH, Shaw GM, Gao F;  
XX WPI; 2000-365651/31.  
XX  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX  
XX  
PS Claim 41; Fig 15; 131pp; English.  
XX  
XX The present in invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 1003 AA;

Alignment Scores:  
Pred. No.: 2,96e-206 Length: 1003  
Score: 3658.00 Matches: 681  
Percent Similarity: 98.61% Conservative: 26  
Best Local Similarity: 94.98% Mismatches: 10

Query Match:	79.82%	Indels:	0
DB:	3	Gaps:	0
US-09-610-313B-30 (1-2469) x AAB69289 (1-1003)			
QY	220	TTCTTCCGCGAGGACTGGCTTCCCTCCAGGCGAAGCGCCGAGTTCCTCCAGCGAGCAG	279
DB	1	PhePheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerSerGln	20
QY	280	AACCGCGCAACAGCCCAACAGCGCGAGCTGCAGGTGCGCGCGCAACAACCCCGCGCAGC	339
DB	21	AlaArgAlaAsnSerProThrSerArgGluLeuGlnValGlnGlyAspAsnProArgSer	40
QY	340	GAGCGCGCGCGAGCGCGAGCGCAACCTGAACTTCCCTCCAGATCACCTGTGGCAGCGC	399
DB	41	GluAlaGlyValGluArgGlnGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg	60
QY	400	CCCTGTGTGACATCAAGGTGGCGGCGAGATCAGGAGGCGCTCTCGACACCGCGCC	459
DB	61	ProLeuValSerIleLysValGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla	80
QY	460	GACGACACCGTGTGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGGC	519
DB	81	AspAspThrValLeuGluGluLeuAlaLeuProGlyArgTrpLysProLysMetIleGly	100
QY	520	GGCATCGCGCGCTTCATCAAGGTGGCGAGTACGACCAAGATCTCTGATCGAGATCTGGCGC	579
DB	101	GlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGly	120
QY	580	AAGAAGGCGCATCGCACCGCTGATCGCGCCCAACCCCGTGAACATCATCTCGCGCAAC	639
DB	121	LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsp	140
QY	640	ATGTGTACCCAGCTGGGTGACCTGAACTTCCCATCGCCCATCGACACCGTGGCC	699
DB	141	MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro	160
QY	700	GTGAAGTGAAGCCCGCATCGACGGCCCAAGGTGAAGTAGTGGCCCTGACCGAGAG	759
DB	161	VallysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu	180
QY	760	AAGATCAAGGCGCTGACCGCATCGCAGGAGATGGAGAGGAGGCGCAAGATCACCAAG	819
DB	181	LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys	200
QY	820	ATCGGCGCGGAGAACCCCTACAAACCCCGTGTTCGCCATCAAGAAGAAGACAGACAC	879
DB	201	IleGlyProGluAsnProTyAsnThrProValPheAlaIleLysArgLysAspSerThr	220
QY	880	AAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGGAG	939
DB	221	LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu	240
QY	940	GTGACGTGGGCATCCCGCCACCGCGCGCTGAAGAAGAAGAAGAGGTGACCGTGTG	999
DB	241	ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu	260
QY	1000	GAGTGGCGGACGCTACTTCAAGCGTGCCTGACGAGGACTTCCGCAAGTACACCGCC	1059
DB	261	AspValGlyAspAlaTyPheSerValProLeuAspGluGlyPheGlyLysTyThrAla	280
QY	1060	TTCAACCATCCCGACATCAACAAGCAGACCCCGCGCATCCGCTACCAAGTCAACGTCGTG	1119
DB	281	PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTrpAsnValLeu	300
QY	1120	CCCAGGCGTGAAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAG	1179
DB	301	ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu	320
QY	1180	CCCTTCCGCGCGCAACCCGAGATGTGATCTACAGTACATGAGCAGCCTGTACGTG	1239
DB	321	ProPheArgAlaArgAsnProLysIleValIleTyGlnTrpMetAspPleutyTyVal	340

QY	1240	GGCAGCGACCTGGAGATCGCGCAGCACCGCGCCCAAGATCGAGGAGCTGGCGAAGCACCTG	1299
DB	341	GlySerAspLeuGluIleGlyHisArgAlaLysIleGluGluLeuArgAlaHisLeu	360
QY	1300	CTGCGCTGGGCTTACCACCCCGACAAAGAACACCAAGAGGAGCCCTTCTCTGTGG	1359
DB	361	LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp	380
QY	1360	ATGGGCTACGAGCTGCACCCCGACCAAGTGGACCTGTGACGCCATCTGAGCTGCCGAGAAG	1419
DB	381	MetGlyTyTrpGluLeuHisProAspLysTrpValGlnProIleLysLeuProGluLys	400
QY	1420	GAGAGCTGGACCGCTGACGACATCCAGAGCTGGTGGCAAGCTGAACTGGCGCCAGCCAG	1479
DB	401	AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln	420
QY	1480	ATCTACCCCGCATCAAGGTGCGCAGCTGTGCAAGCTGTGCGGGCGCCCAAGCCCTG	1539
DB	421	IleTyTrpGlnIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu	440
QY	1540	ACCGACATCTGTGCCCTGTACCGAGAGCGCCGAGCTGGAGCTGGCGGAACCGCGAGATC	1599
DB	441	ThrAspIleValProLeuThrGluAlaGluLeuGluLeuAlaGluAsnArgGluIle	460
QY	1600	CTGCGCGAGCCCGTGCACGGGTGTACTACACCCCGACCAAGAGACTGTGTGGCGGAGATC	1659
DB	461	LeuLysGluProValHisGlyValTyTrpAspProSerLysAspLeuIleAlaGluIle	480
QY	1660	CAGAAGCAGGCGCACGACGAGTGCACCTTACCAGATCTACAGGAGCCCTTCAAGAACCTG	1719
DB	481	GlnLysGlnGlyHisAspGlnTrpTrpGlnIleTyTrpGlnGluProPheLysAsnLeu	500
QY	1720	AAGACCGCGCAAGTACGCGCAAGATCGCACCCCGCCACACACGACGCTGAAGCAGCTGACC	1779
DB	501	LysThrGlyLysTyAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr	520
QY	1780	GAGCGCTGCAAGAAGATCGCCATCGAGAGCATCTGTGATCTGGGGCAAGACCCCGAGTTC	1839
DB	521	GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp**LysThrProLysPhe	540
QY	1840	CGCTGCGCCATCCAGAAGGAGAGCTGGGAGACCTGGTGGACCGACTACTGCGACGCCACC	1899
DB	541	ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyTrpGlnAlaThr	560
QY	1900	TGGATCCCGCAGTGGGAGTTCGTGAACACCCCGCCCTGTGTGAAGCTGTGTATCCAGCTG	1959
DB	561	TrpIleProAspTrpGluPheValAsnThrProProLeuValLysLeuTrpTyTrpGlnLeu	580
QY	1960	GAGAAGGAGCCATCATCTCGCGCGCAGACCTTCTACGTGAGCGCGCGCCCAACCGCGAG	2019
DB	581	GluLysGluProIleValGlyAlaGluThrPheTyValAspGlyAlaAlaAsnArgGlu	600
QY	2020	ACCAAGATCGCAGGCGCGCTAGTACCGACCGGGCGCGCAGAGATCGTGGAGCTG	2079
DB	601	ThrLysValGlyLysAlaGlyTyTrpValThrAspArgGlyArgGlnLysIleValSerLeu	620
QY	2080	ACCGAGACCAACCAAGAGAGCCAGCTCGAGGCCATCCAGCTGGCGCCCTCGACGACAGC	2139
DB	621	ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer	640
QY	2140	GGCAGCGAGGTGAACATCTGTGACCGCAGCAGCTACCGCTTGGGCATCATCCAGGCCCG	2199
DB	641	GlyThrGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIleIleGlnAlaGln	660
QY	2200	CCCGAAGAGCGGAGGAGCTGTGAACAGATCATCGAGCTGAGCTGATCAAGAGGAG	2259
DB	661	ProAspLysSerGluSerGluLeuValAsnGlnIleIleGlnLeuLeuIleAsnLysGlu	680
QY	2260	AAGTGTACCTGAGCTGGTGGCGCCGACCAAGGGCATCGCGCGCAACGAGCAGATCGAC	2319
DB	681	ArgValTyTrpLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnGluValAsp	700
QY	2320	AAGCTGTGAGCAAGGCGCATCCGCAAGGTGCTGTCTCTGAGCGGCATCGAT	2370



Db 421 IleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
Qy 1540 ACCGACATCGTGGCCCTGACCGAGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATC 1599  
Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuLeuAlaGluAsnLysGluIle 460  
Qy 1600 CTGCGCGAGCCGCTGCGCGCTGTACTACGACCCAGCAGGACCTGGTGGCGAGATC 1659  
Db 461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480  
Qy 1660 CAGAAGCAGGCGCCACGACCTGACCTACAGATCTACAGGAGCCCTTCAAGAACCTG 1719  
Db 481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500  
Qy 1720 AAGCCGCGAAGTACGCGAAGATCGCACCGCCACACGACGCTGAAGCAGCTGACC 1779  
Db 501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520  
Qy 1780 GAGCGCTGCAGAGATCGCCATCGGAGAGCATCTGATCTGGGCGAAGACCCCAAGTTC 1839  
Db 521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTrpGlyLysIleProLysPhe 540  
Qy 1840 CGCTGCGCCATCCAGAAGGACCTGGGAGACCTGGTGACCGACTACTGGCGAGCCAC 1899  
Db 541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAlaThr 560  
Qy 1900 TGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGTACCACTG 1959  
Db 561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTrpTyrGlnLeu 580  
Qy 1960 GAGAAGGAGCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCCCAACCGCGAG 2019  
Db 581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600  
Qy 2020 ACCAAGATCGCAGGCGCGCTAGTGCACGACCGCGCGCGAGAGATCGTGAGCTG 2079  
Db 601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620  
Qy 2080 ACCGAGACCCACCAACGAGACCGAGCTGCAGGCGCCATCCAGCTGGCGCTGCAGGACAGC 2139  
Db 621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640  
Qy 2140 GCGAGCAGGTGAATCATGTCAGCAGCAGCAGCTAGCGCTGGCGCATCATCCAGGCCAG 2199  
Db 641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660  
Qy 2200 CCGCACAAGAGCAGAGAGCTGTGTGAACACGATCATCGAGCAGCTGATCAAGAAGGAG 2259  
Db 661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysGlu 680  
Qy 2260 AAGGTGTACCTGAGCTGGGTGCCCGCCACCAAGGGCATCGCGCGCAACGAGCAGATCGAC 2319  
Db 681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAsp 700  
Qy 2320 AAGCTGTGACCAAGGGCATCCGCAAGGTGCTGTCTCTGAGCGGCATCGAT 2370  
Db 701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 6  
AAB69287  
ID AAB69287 standard; protein; 1005 AA.  
AC AAB69287;  
XX  
XX  
DT 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX  
XX HIV-1 non-subtype B clone 96ZM751-3 pol protein.  
XX  
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
XX vif; vpr; tat; rev; nef; vaccine.  
XX  
OS Human immunodeficiency virus 1.

XX WO200026416-A1.  
XX 11-MAY-2000.  
XX 25-OCT-1999; 99WO-US024837.  
XX 02-NOV-1998; 98US-00184418.  
XX (UABR-) UAB RES FOUND.  
XX Hahn BH, Shaw GM, Gao F;  
XX WPI; 2000-365651/31.  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
XX type 1 useful for detecting and treating AIDS comprises a specific  
XX nucleotide sequence.  
XX Claim 41; Fig 15; 13lpp; English.  
XX The present invention provides the protein and coding sequences for a  
XX number of human immunodeficiency virus (HIV) type 1 non-subtype B  
XX isolates. The sequences shown include the near full-length coding  
XX sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
XX rev and nef proteins. These can be used to detect the presence of HIV-1  
XX in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
XX These antibodies can be used in vaccines to prevent and treat HIV  
XX infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX  
SQ Sequence 1005 AA;  
Alignment Scores:  
Pred. No.: 1.96e-205 Length: 1005  
Score: 3644.00 Matches: 685  
Percent Similarity: 97.23% Conservative: 18  
Best Local Similarity: 94.74% Mismatches: 14  
Query Match: 79.51% Indels: 6  
DB: 3 Gaps: 1  
US-09-610-313B-30 (1-2469) x AAB69287 (1-1005)  
Qy 220 TTCTTCGCGAGACCTGGCTTCTCCCGAGGCGAGCCCGAGTGTCCCGAGCGAGCAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20  
Qy 280 AACCGCGCC-----AACAGCCCCACCGCGCGAGCTCGAGTGTGCGC 321  
Db 21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40  
Qy 322 GCGGACAAACCCCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 381  
Db 41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60  
Qy 382 ATCACTGTGGCAGCGCCCTGGTGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCGCG 441  
Db 61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIleLysGluAla 80  
Qy 442 CTGCTGGACACCG 501  
Db 81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluIleAsnLeuProGlyLysTrp 100  
Qy 502 AAGCCCAAGATGATCGCGCGCGCATCGCGCGCTTCAAGGTGCGCGCGCGCGCGCGCG 561  
Db 101 LysProLysMetIleGlyGlyGlyGlyPheIleLysValArgGlnTyrAspGlnIle 120  
Qy 562 CTGATCAGATCTGCGCGCGCAAGGCCATCGGCAACCGTGTGTGATCGCGCGCGCGCG 621  
Db 121 LeuIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyProThrProVal 140  
Qy 622 AACATCATCGCGCGCAACATGCTGACCGCGCTGGCTGCGCGCGCGCGCGCGCGCGCG 681  
Db 141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160

QY 682 CCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGCATGGACCGGCCCAAGGTGAAGCAG 741  
 DB 161 ProileGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln 180  
 QY 742 TGCCCTGTGACGAGGAGAAATCAAGGCCCTGACCGCCATCTGCAGAGAGATGGAAG 801  
 DB 181 TriProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys 200  
 QY 802 GAGGGCAAGATCAACAGATCGGCCCGGAGAACCCCTACACACCCCGTTCGCCATC 861  
 DB 201 GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle 220  
 QY 862 AAGAAGAGACAGACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTTGAACAAGCGC 921  
 DB 221 LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArg 240  
 QY 922 ACCAGGACTTCTGGAGGTGACGTGGGATCCCGCATCCCGCCCGCCCTGAAGAAG 981  
 DB 241 ThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLys 260  
 QY 982 AAGAGGTGACCGTGTGGAGTGGGAGCGCCCTACTTACGCTGGCCCTGGACGAGGAC 1041  
 DB 261 LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly 280  
 QY 1042 TTCGCGAAGTACACCGCTTCAACATCCCGCATCAACCAACGAGACCCCGCCGATCCGC 1101  
 DB 281 PheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArg 300  
 QY 1102 TACAGTACAACTGTGTCGCCCGGCTGGAAGGGCAGCGCCAGCATCTTCCAGAGCAGC 1161  
 DB 301 TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer 320  
 QY 1162 ATGACCAAGATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATCTGATCTACCACTAC 1221  
 DB 321 MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr 340  
 QY 1222 ATGAGCACTGTACGTGGGCGAGCAGCTGGAGATCGGCAGCAGCCGCGCAAGATCGAG 1281  
 DB 341 MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu 360  
 QY 1282 GAGCTGCGCAAGCACTGTGCTGCGCTGGGGTTCCACACCCCGCAAGACCAAGCAAG 1341  
 DB 361 GluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLys 380  
 QY 1342 GAGCCCTCTCTGTGGATGGGTACGAGCTGCACCGCCCGCAAGTGGACCGTGCAGCCC 1401  
 DB 381 GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPro 400  
 QY 1402 ATGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAAG 1461  
 DB 401 IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys 420  
 QY 1462 CTGAACCTGGCCGACCATCTACCCCGCATCAAGGTGGCCGAGCTGTCAAGCTGCTG 1521  
 DB 421 LeuAsnTrp\*\*SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu 440  
 QY 1522 CGCGGCGCAAGCCCTGACGACATCGTCCCTGACCGAGGCGCGAGCTGGAGCTG 1581  
 DB 441 ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeu 460  
 QY 1582 GCCGAGAACCGGAGATCTTGGCGGAGCCCGTGCACCGCTGTACTACGACCCCGCAAG 1641  
 DB 461 AlaGluSerArgGluIleLysLysGluProValHisGlyValTyrTyrAspProSerLys 480  
 QY 1642 GACTGTGTGGCCGAGATCCAGAAGACGAGGCGCACGACAGTGGACCTACAGATCTACCA 1701  
 DB 481 AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln 500  
 QY 1702 GAGCCCTTCAAGACCTGAAGACCGGCAAGTACCGCAAGATGCGCACCGCCCAACCAAC 1761  
 DB 501 GluProPheLysAsnLeuLysThrGlyLysThrAlaLysMetArgThrAlaHisAsn 520

QY 1762 GACGTGAACAGCAGCTGATCCAGGCGCGTGCAGAGATCGCCATGAGAGCATCTGTGATCTGG 1821  
 DB 521 AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrp 540  
 QY 1822 GGCAGACCCCAAGTTCCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGCTGGAC 1881  
 DB 541 GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr 560  
 QY 1882 GACTACTGCGAGCCACCTCGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTGGTG 1941  
 DB 561 AspTyrTrpGluAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal 580  
 QY 1942 AAGCTGTGTACAGCTGGAGAGGAGCCATCATCGGCCCGAGACCTTCTACGTGGAC 2001  
 DB 581 LysLeuTyrTrpGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp 600  
 QY 2002 GCGCGCCCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTGACCGACCGGCGCG 2061  
 DB 601 GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg 620  
 QY 2062 CAGAAGATCTGTGAGCTGACCGAGACCCCAACCAAGACCGAGCTGAGGCGCATCCAG 2121  
 DB 621 GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln 640  
 QY 2122 CTGGCCCTGCGAGACAGCGGAGGTCAACATCGTACCGACAGCGAGTACGCCCTG 2181  
 DB 641 LeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeu 660  
 QY 2182 GGCATCATTCAGGCCCGCGAGCGGACAGAGCGGAGCTGGTGAACCAAGCATCTCGAG 2241  
 DB 661 GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGlu 680  
 QY 2242 CAGCTCATCAAGAGAGAGAGTGTACTGAGCTGGGTGCCCGCCACCAAGGCGCATCGGC 2301  
 DB 681 GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly 700  
 QY 2302 GGCAGCAGCAGATCCACAGCTGGTGTGAGCAAGGCAATCCGCAAGGTGCTGTCTGGAC 2361  
 DB 701 GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp 720  
 QY 2362 GGCATCGAT 2370  
 DB 721 GlyIleAsp 723  
 RESULT 7  
 AAE37601  
 ID AAE37601 standard; protein; 998 AA.  
 XX AAE37601;  
 AC AAE37601;  
 XX  
 DT 23-OCT-2003 (revised)  
 DT 27-AUG-2003 (first entry)  
 XX  
 DE HIV-1 subtype C isolate Dul51 reverse transcriptase (RT) protein.  
 XX  
 KW Regulatory gene; accessory gene; HIV; human immunodeficiency virus;  
 KW vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN WO2003037919-A2.  
 XX  
 XX 08-MAY-2003.  
 XX  
 PF 31-OCT-2002; 2002WO-1B004550.  
 XX  
 PR 31-OCT-2001; 2001ZA-00008978.  
 XX  
 PA (SAME-) SOUTH AFRICAN MEDICAL RES. COUNCIL.  
 PA (UYCA-) UNIV CAPE TOWN.  
 XX  
 PI Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;  
 XX







QY 723 CGGCCCCAAGGTGAAGCAAGTGGCCCTTACCGAGGAGAGATCAAGGCCCTTGACGGCCAT 782  
 Db 724 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIle 674  
 QY 783 CTGGAGGAGATGGAGAGAGGAGGAGATCACCAAGATCGGCCCGGAGAACCCCTACAA 842  
 Db 784 eCyAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAs 694  
 QY 843 CACCCCGTGTTCGCCATCAAGAAGAGGAGCAGCAGCAAGTGGCGCAAGCTGGTGACTT 902  
 Db 844 nThrProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPh 714  
 QY 903 CCGGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCGCCACC 962  
 Db 904 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734  
 QY 963 CGCCGGCTGAAGAAGAGAGAGCGTGACCGTGTGGAGCTGGCGAGCGCTACTTCCAG 1022  
 Db 964 oAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754  
 QY 1023 CGTCCCGCTGGAGGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAA 1082  
 Db 1024 rIleProLeuTyrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774  
 QY 1083 CGAGACCCCGGATCCGCTACCAAGTACACCGTGTGCGCCCGAGGCTGGAGAGGCGAGCCC 1142  
 Db 1084 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerLe 794  
 QY 1143 CAGCATCTCCAGAGCAGCATGACCAAGATCCTCGAGCCCTTCCGCGCCCGCAACCCGA 1202  
 Db 1144 uAlaIlePheGlnSerMetThrLysThrLysLeuGluProPheArgLysGlnAsnProGl 814  
 QY 1203 GATCGTATCTACCAAGTACAGGAGCTGACGTCGAGGAGGAGCAGCTGGAGATCGGCCA 1262  
 Db 1204 yIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuLysIleGlyGl 834  
 QY 1263 GCACCGGCCAAGATCGAGGAGTGCACAGCACCTGCTGCGTGGGCTTCCACCCGCC 1322  
 Db 1264 nHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrPr 854  
 QY 1323 CGACAGAGCACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1382  
 Db 1324 oAspLysLysHis---LysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAs 873  
 QY 1383 CAAGTGACCGTGCAGCGCCATCGAGCTGCCGAGAGGAGAGGAGGAGGAGGAGGAGGAG 1442  
 Db 1384 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893  
 QY 1443 CCAGAAGCTGGTGGCAAGCTGAATCGGCGAGCCAGATCTACCCCGCATCAAGGTGCG 1502  
 Db 1444 eGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValAr 913  
 QY 1503 CCAGCTGTGACAGCTGTGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1562  
 Db 1504 gGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933  
 QY 1563 GGAGCGGAGCTGGCGGAGAGAACCGGAGATCTGCGGAGCGCGGAGCGGAGCGGAGCG 1622  
 Db 1564 uGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysGlnProValHisGlyVa 953  
 QY 1623 GTACTACGACCCAGCAGGAGCTGTGGCGAGATCCAGAAGCAGGCGGAGGAGGAGGAGG 1682  
 Db 1624 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973  
 QY 1683 GACCTACAGATCTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1742  
 Db 1684 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993  
 QY 1743 GCGCAGCGCCACACCAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1802  
 Db 1744 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

QY 1803 GGAGAGCATCGTGTGATCTGGGGCAAGACCCCAAGTTCGCTGCGCATCCAGAGAGAC 1862  
 Db 1804 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033  
 QY 1863 CTGGAGACCTGGTGGAGCCGACTACTGGCAGGCACCTGGATCCCGAGTGGAGTTCGT 1922  
 Db 1864 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053  
 QY 1923 GAACACCCCGCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGGAGGAGGAGGAGGAG 1982  
 Db 1924 lAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysAspProIleValGlyVa 1073  
 QY 1983 CGAGACCTTCTACCTGTGAGCGCGCCCAACCGCGAGACCAAGATCCGCGAGGCGCGCTA 2042  
 Db 1984 lGluThrPheTyrValAspGlyAlaAlaAsnArgLysThrLysIleGlyLysAlaGlyTy 1093  
 QY 2043 CGTACCGACCGCGCGCGAGAGATCGTGAGCTGACCGAGACCAACCAACCAAGAGAC 2102  
 Db 2044 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113  
 QY 2103 CGAGCTCGAGGCCATCAGCTGGCCCTGCGAGCAGCGGCGAGCGGAGTGAACATCGTGAC 2162  
 Db 2104 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133  
 QY 2163 CGACAGCAGTACGCCCTGGGCATCATCCAGGCCCGCCCGCCCGAGAGCGGAGCGCT 2222  
 Db 2164 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153  
 QY 2223 GGTCAACACAGATCATCAGCAGCTGATCAAGAGGAGAGGAGTGTACCTGAGCTGGGTGCC 2282  
 Db 2224 uValAsnGlnIleIleGluGlnLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173  
 QY 2283 CGCCCAACAGGCGATCGCGCGCAACCGAGCAGATCGACAAGCTGGTGAGCAAGGCGCATCG 2342  
 Db 2284 oAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleAr 1193  
 QY 2343 CAAGTGTCTGTCTGACCGCGCATCGAT 2370  
 Db 2344 gLysValLeuPheLeuAspGlyIleAsp 1202  
 RESULT 9  
 AAB69282  
 ID AAB69282 standard; protein; 1000 AA.  
 AC AAB69282;  
 XX  
 XX 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 XX HIV-1 non-subtype B clone 92RW009-6 pol protein.  
 DE HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 XX Human immunodeficiency virus 1.  
 OS  
 FN WO200026416-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 25-OCT-1999; 99WO-US024837.  
 XX  
 PR 02-NOV-1998; 98US-00184418.  
 XX  
 PA (UABR-) UAB RES FOUND.  
 XX  
 PI Hahn BH, Shaw GM, Gao F;  
 XX  
 DR WPI; 2000-365651/31.  
 XX  
 PT Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.

XX Claim 41; Fig 15; 131pp; English.

XX The present in invention provides the protein and coding sequences for a

CC number of human immunodeficiency virus (HIV) type 1 non-subtype B

CC isolates. The sequences shown include the near full-length coding

CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,

CC rev and nef proteins. These can be used to detect the presence of HIV-1

CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.

CC These antibodies can be used in vaccines to prevent and treat HIV

CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 1000 AA;

SQ

Alignment Scores:

Pred. No.:	3 07e-200	Length:	1000
Score:	3555.50	Matches:	664
Percent Similarity:	96.52%	Conservative:	29
Best Local Similarity:	92.48%	Mismatches:	24
Query Match:	77.58%	Indels:	1
DB:	3	Gaps:	1

US-09-610-313B-30 (1-2469) x AAB69282 (1-1000)

QY 220 TTCTTCCGAGACCTGGCTTCCCTCCAGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279

DB 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20

QY 280 AACCGGCGCAAGCCCCACCACCGCGGAGCTCAGGTGGCGCGC---GACAACCCCGCG 336

DB 21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyArgAspSerLeuSer 40

QY 337 AGCGAGCGCGCGCGAGCGCCAGGCGCACCTCAACTTCCCTCCAGATCACCTGTGGCAG 396

DB 41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnLeuLeuTrpGln 60

QY 397 CGCCCTCGTGATCAAGTGGCGCGCCAGATCAAGAGAGCCCTGTGGACACCGCG 456

DB 61 ArgProLeuValThrValIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80

QY 457 GCGAGACACCTGTGTGGAGGAGATGAGCTCCCGCGCAAGTGGAGCCCAAGATGATC 516

DB 81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIle 100

QY 517 GCGGAGATCGCGGCTTCATCAAGTGGCGCGCTAGCACAGATCCTGTGATCGAGATCTGC 576

DB 101 GlyGlyIleGlyPheIleIleValLysGlnIleThrAspGlnLeuLeuIleGluIleCys 120

QY 577 GCGAAGAGCCATCGGCACCGCTGTGATCGGCGCCACCCCGTGAAACATCATCGCGCGC 636

DB 121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleIleGlyArg 140

QY 637 AACATGTGACCGAGCTGGCTGACCTGAACTTCCCATAGCCCGCCATCGAGACCGTG 696

DB 141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160

QY 697 CCCTGAAGTGAAGCCCGCATGGAGCGGCCCAAGTGAAGCAGTGGCCCTGACCGAG 756

DB 161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 180

QY 757 GAGAAGATCAAGCCCTTGACCGGCATCTGCGAGGAGATGGAGAAGGAGGCGCAAGATCAC 816

DB 181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluLysLysIleSer 200

QY 817 AGATCGGCCCGGAGAACCCCTTCAACACCCCGCTGTTCGCCATCAAGAAAGAGGACAGC 876

DB 201 LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 220

QY 877 ACCAAGTGGCGAGCTGGTGGACTTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGG 936

DB 221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240

QY 937 GAGGTGCAGCTGGGCGATCCCCCAACCCCGCGCTGGAAGAAGAAGAGCGTGCAGCGTG 996

DB 241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVal 260

QY 997 CTGGAGCTGGGCGACCGCCTTACGCTGCGCCCTCGAGGAGACTTCCGCAAGTAGTACAC 1056

DB 261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280

QY 1057 GCCTTCCACCATCCCGAGCATCAACAAGAGACCCCGGATCCGCTACCAAGTCAACAGTG 1116

DB 281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300

QY 1117 CTGCCCCAGGCTGGAGAGGCGACCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTG 1176

DB 301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320

QY 1177 GAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATCTACCAAGTACATGAGACCGCTGTAC 1236

DB 321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspLeuTyr 340

QY 1237 GTGGGAGCGACCTGGAGATCGGCCAGCACCGGCCCAAGATCGAGAGCTGCCAAGCAC 1296

DB 341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHis 360

QY 1297 CTGCTCGCTGGGCTTACCACCCCGCAAGAACACCAAGAGAGCCCGCTTCTCTG 1356

DB 361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 380

QY 1357 TGGATCGGCTACGAGCTGCACCCGCAAGTGCAGCTGCAGCCCATCGAGCTGCCCGAG 1416

DB 381 TrpMetGlyTyrGlnLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400

QY 1417 AAGGAGAGCTGGACCGTGAAACGATCCAGAACCTGCTGGGCAAGCTGAACTGGGCGACG 1476

DB 401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420

QY 1477 CAGATCTACCCCGCATCAAGTGGCGCGCAGCTGTGCAAGCTGCTGCGCGCGCCAGGCC 1536

DB 421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440

QY 1537 CTGACCGACATCTGCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGAGAACCGCGAG 1596

DB 441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 460

QY 1597 ATCTCTCGCGAGCCGCTGCACGCGCTGTACTACGACCCCAAGAGCAGCTGTGGCGCGAG 1656

DB 461 IleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGlu 480

QY 1657 ATCCAGAGAGGCGCCACGACAGTGCAGCTACAGATCTACAGAGCGCTTCAAGAAC 1716

DB 481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500

QY 1717 CTGAAGACCGCAAGTACGCAAGATGCGCACCGCCACACCAACGACGCTGAAGCAGCTG 1776

DB 501 LeuLysThrGlyLysTyrAlaLysArgThrAlaHisThrAsnAspValLysGlnLeu 520

QY 1777 ACCGAGCGCTGCAGAAGATCGCCATCGAGAGCATCTGTATCTGGGCAAGACCCCAAG 1836

DB 521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleIleTrpGlyLysThrProLys 540

QY 1837 TTCCGCTGCGCATCCAGAGGAGACCTGGGAGACCTGTGGAGCCGACTACTGCGAGGCC 1896

DB 541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAla 560

QY 1897 ACCTGGATCCCGAGTGGAGTTCTGTGAACACCCCGCTGTGTGAAGCTGTGTGTACCG 1956

DB 561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580

QY 1957 CTGGAAGAAGGAGCCCATCATCTCGCGCGCGAGACCTTCTACGTGAGCGCGCGCAACCGC 2016

DB 581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

QY 2017 GAGACCAAGTCCGCAAGCGCGCTACGTGACCGACCGGCGCGCGAGAGATCGTGAGC 2076

Db 601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620  
 Qy 2077 CTGACCGAGACCAACAGAGAGAGCTGCAGGCGCCATCCAGCTGGCCCTCGAGGAC 2136  
 Db 621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640  
 Qy 2137 AGCGGACGAGGTGAACATCGTACCGACAGCCAGGCTAGCGCCCTGGGCATCATCCAGGCC 2196  
 Db 641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleGlnAla 660  
 Qy 2197 CAGCCGACAGAGAGCGAGCGAGCTGGTGAACAGAGATCATCGAGCAGCTCATCAAGAAG 2256  
 Db 661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleGlnLeuLysLys 680  
 Qy 2257 GAGAAGGTGTACTGAGCTGGTGGTCCGCCACAGGCGCATCGCGGCAACGAGCAGATC 2316  
 Db 681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700  
 Qy 2317 GACAAGCTGTGAGCAAGGCGATCGGCAAGGTGCTGTCTCGAGCGGCGATCGAT 2370  
 Db 701 AspLysLeuValSerSerGlyIleArgArgValLeuPheLeuAspGlyIleAsp 718  
 RESULT 10  
 AAW72993  
 ID AAW72993 standard; protein; 1002 AA.  
 XX AC AAW72993;  
 XX DT 25-MAR-2003 (revised)  
 XX DT 15-FEB-1999 (first entry)  
 XX DE HIV isolate LAV.MAL pol protein.  
 XX KW LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;  
 XX KW pol.  
 XX OS Human T-lymphotropic virus.  
 XX PN US5824482-A.  
 XX PD 20-OCT-1998.  
 XX PF 06-JUN-1995; 95US-00471474.  
 XX PR 23-JUN-1986; 86FR-00040138.  
 XX PR 13-APR-1987; 87US-00038330.  
 XX PR 19-FEB-1991; 91US-00656797.  
 XX PR 10-DEC-1992; 92US-00988530.  
 XX PR 18-NOV-1993; 93US-00154397.  
 XX PA (INSP ) INST PASTEUR.  
 XX PI Wain-Hobson S, Sonigo P, Alizon M, Montagnier L;  
 XX DR WPI; 1998-582548/49.  
 XX DR N-PSDB; AAV63467.  
 XX PT Human immunodeficiency virus isolate LAV(MAL) - and method for detecting  
 XX PT anti-HIV antibodies.  
 XX PS Disclosure; Fig 7B-E; 47pp; English.  
 XX CC This is the amino acid sequence of the pol protein of lymphadenopathy-  
 CC associated virus LAV.MAL (CNCM I-641), a new virus isolate from Zaïre  
 CC that is responsible for diseases clinically related to AIDS. The sequence  
 CC was deduced from an open reading frame (ORF) of the LAV.MAL genome (see  
 CC AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of  
 CC the envelope glycoprotein can be used as antigens in a claimed method for  
 CC the in vitro detection of an antibody directed against LAV. The method is  
 CC useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in  
 CC patients, asymptomatic carriers and in blood-related products. LAV viral  
 CC antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct  
 CC PR field.)

XX SQ Sequence 1002 AA;  
 Alignment Scores:  
 Pred. No.: 3,49e-199 Length: 1002  
 Score: 3537.50 Matches: 657  
 Percent Similarity: 96.67% Conservatives: 39  
 Best Local Similarity: 91.25% Mismatches: 21  
 Query Match: 77.19% Indels: 3  
 Gaps: 2  
 DB:  
 US-09-610-313B-30 (1-2469) x AAW72993 (1-1002)  
 Qy 220 TTCTTCCGAGAGACCTGGCTTCCCGAGGCAAGGCCCGCGAGTTCCCGAGGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGlnGln 20  
 Qy 280 AACCGCGCCACAGCCCGCCAGCGCGGAGCTGCAGGTG---CGCGGCGCAACCCCGCGC 336  
 Db 21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40  
 Qy 337 AGCGAGGCGCGCGCGAGCGCGAGGCG---ACCCTGAACCTTCCCGAGATCACCCCTG 390  
 Db 41 SerGluThrGlyAlaGluArgGlnGlyIleValSerPheSerPheProGlnIleThrLeu 60  
 Qy 391 TGGCAGCGCCCTGGTGAGCATCAAGGTGGCGGCCAGATCAAGAGGCGCCCTGCTGGAC 450  
 Db 61 TrpGlnArgProValValThrValArgValGlyGlyGlnLeuLysGluAlaLeuAsp 80  
 Qy 451 ACCGCGCGCCAGCAGCAGCTGCTGGAGGAGATGAGCTGCCGCGGCAAGTGAAGCCCAAG 510  
 Db 81 ThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLys 100  
 Qy 511 ATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGCCAGTACGACCATCTGATCCAG 570  
 Db 101 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120  
 Qy 571 ATCTGCGCGCAAGAGCCATCGGCACCGTGTGATCGGCCCGCCACCCCGTGAACATCATC 630  
 Db 121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140  
 Qy 631 GCGCGCAACATGCTGACCCAGCTGGCTGGCGCCACTTCCCATCAGCCCATCCAG 690  
 Db 141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
 Qy 691 ACCGTGCGCGTGAAGCTGAAGCCCGCATGACCGCGCCCAAGGTGAAGTGAAGTGGCCCTG 750  
 Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180  
 Qy 751 ACCGAGGAGAAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGAGGGCAAG 810  
 Db 181 ThrGluLysLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200  
 Qy 811 ATCAACCAAGATCGGCCCGCGAGACCCCTTACACACCCCGCTGTCGCGCATCAAGAGAG 870  
 Db 201 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 220  
 Qy 871 GACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGAC 930  
 Db 221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240  
 Qy 931 TTCTGGGAGGTGACGCTGGGCATCCCCCACCCCGCGCTTGAAGAAGAAGAGAGCGGTG 990  
 Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
 Qy 991 ACCGTGCTGACGTGGCGGAGCCCTACTTACGCGTGGCCCTTGGAGCAGGAGCTTCGCGAAG 1050  
 Db 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 280  
 Qy 1051 TACACCGCCCTTACCATCCCGCAGCATCAACAAGCAGACCCCGCGCATCCGCTACAGTAC 1110  
 Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyLysIleArgTyrGlnTyr 300



## Alignment Scores:

Pred. No.: 3.59e-199 Length: 854  
Score: 3537.00 Matches: 667  
Percent Similarity: 94.44% Conservative: 30  
Best Local Similarity: 90.38% Mismatches: 30  
Query Match: 77.18% Indels: 11  
DB: 5 Gaps: 2

US-09-610-313B-30 (1-2469) x AAM48949 (1-854)

```
QY 171 CGGCAAGGAGGGCCACACAGATGAAGGACTGCACCGAGCGCCAGGCGCAACTCTT----- 224
DB 118 ArgGlnGlyPheProSerHisAspValValLysArgArgProValProSerLeuHisAla 137
QY 225 -----CCGCGAGGACTTGCC-----TTCCCCCAAGGCAAGGCCCG 260
DB 138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAlaAr 157
QY 261 CGAGTTCCTCCAGCGAGCAGAAACCGCGCCAAACAGCCCAACAGCCCGAGCTGCAGGTGCG 320
DB 157 gGlnPheProSerGluInThrArgAlaAsnSerProThrSerArgGluLeuGlnValAr 177
QY 321 CGGCGAACCCTCCGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
DB 177 gArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGl 197
QY 381 GATCACCTGTGGCAGCGCCCTCTGTGTGAGCATCAAGGTGGCGCGCGCGCGCGCGCGCGCGCG 440
DB 197 nileThrLeuTrpGlnArgProLeuValSerileLysileGlyGlnThrArgGluAl 217
QY 441 CCTGCTGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
DB 217 aLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuProGlyLysTr 237
QY 501 GAACCCCAAGATGATCGCGCGCATCGCGCGCTTATCAAGTGGCGCGCGCGCGCGCGCGCGCG 560
DB 237 pLysProLysMetileGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIl 257
QY 561 CCTGATCGAGATCTCGCGCAAGAGGCATCGGCACCGTGTGATCGCGCGCGCGCGCGCGCGCG 620
DB 257 eLeuileGluileCysGlyLysLysAlaileGlyThrValLeuValGlyProThrProva 277
QY 621 GAACATCATCGCGCGCAACATGTGACCCAGCTGGCTGCGACCCCTGAACTTCCCGCATCAG 680
DB 277 lAsnileileGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSe 297
QY 681 CCCCATCGAGACCTGCGCGTGAAGCTGAAGCCCGCGCATGACCGCCCAAGGTGAAGCA 740
DB 297 rProileGluThrValProValLysLeuLysPheProGlyMetAspGlyProLysValLysGl 317
QY 741 GTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAA 800
DB 317 nTrpProleuThrGluValLysIleLysAlaLeuThrAlaileCysGluGluMetGluLy 337
QY 801 GGAGGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
DB 337 sGluGlyLysileThrLysileGlyProGluAsnProfyrAsnThrProIlePheAlaIl 357
QY 861 CAAGAAGAGACAGCACCAGTGGCGGCAAGCTGGTGACTTCCGCGAGCTGAACAAGCG 920
DB 357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysAr 377
QY 921 CACCCAGGACTTCGCGAGGTGAGCTGGGCATCCCCACCCCGCGCGCGCGCGCGCGCGCGCGCG 980
DB 377 gThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLy 397
QY 981 GAAGAGCGTCAACCGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040
DB 397 sLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluGl 417
QY 1041 CTTCCGCAAGTACACCGCTTTCACATCCCGCATCAACCGAGACCCCGCGCGCGCGCGCGCG 1100
```

```
DB 417 yPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleAr 437
QY 1101 CTACAGGTACAACGTGTGCTGCCCGGAGGCTGAAGGCGAGCGCCAGCATCTTCCAGAGCAG 1160
DB 437 gTrpGlnTrpAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSe 457
QY 1161 CATGACCAAGATCTCTGAGAGCCCTTCCGGCGCGCGCAACCCCGAGATCGTGATCTACAGTA 1220
DB 457 rMetThrLysIleLeuGluProPheArgAlaLysAsnProGluIleValIleTrpGlnTy 477
QY 1221 CATGCGACGCTGTACGTGGCGACGCTCGAGATCGCGCAGCAGCCGCGCAGCAGCCGCAAGTGA 1280
DB 477 rMetAlaAlaLeuTrpValGlySerAspLeuGluileGlyGlnHisArgAlaLysIleGl 497
QY 1281 GGAGCTCGCGCAAGCACCTGTGCTGGGCTTCCACACCCCGCGCGCGCGCGCGCGCGCGCG 1340
DB 497 uGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLy 517
QY 1341 GGAGCCCGCTTCTCTGTGATGGCTACGAGCTGCACCCCGCGCGCGCGCGCGCGCGCGCGCG 1400
DB 517 sGluProProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnPr 537
QY 1401 CATGAGCTGCCCGCGAGAGGAGAGCTGCACCGTGAACGACATCCAGAAAGCTGGTGGGCAA 1460
DB 537 oileGlnLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLy 557
QY 1461 GCTGAACCTGGCGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGCGCGCGCGCGCGCG 1520
DB 557 sLeuAsnTrpThrSerGlnileTrpProGlyLysValArgGlnLeuCysLysLeuLe 577
QY 1521 GCGCGCGCGCAAGCCCTGACCGATCGTGCCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCG 1580
DB 577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLe 597
QY 1581 GGCGCGAAGCCGCGAGATCTCGCGCGAGCGCGTGCACCGCGTGTACTACGACCCCGCGCAA 1640
DB 597 uAlaGluAsnArgGluileLeuLysGluProValHisGlyValTrpTrpAspProserLy 617
QY 1641 GGACTGCTGCTGCCGAGATCCAGAAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1700
DB 617 sAspLeuileAlaGluileGlnLysGlnGlyAspAspGlnTrpThrTrpGlnileTrpGl 637
QY 1701 GGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCCAAGATGCGCACCGCGCGCGCGCG 1760
DB 637 nGluProPheLysAsnLeuLysThrGlyLysTrpAlaLysArgArgThrThrHisThrAs 657
QY 1761 CGAGCTGACGAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1820
DB 657 nAspValLysGlnLeuThrGluAlaValGlnLysileSerLeuGluSerIleValThrTr 677
QY 1821 GGCGAAGACCCCGCAAGTTCGCGCTGCCCATCCAGAAAGAGAGACCTGGGAGACCTGGTGAC 1880
DB 677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluileTrpTrpTh 697
QY 1881 CGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCGCGCGCGCG 1940
DB 697 rAspTrpTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVa 717
QY 1941 GAAGCTGTGCTGACGCTGGAGGAGGCCCATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2000
DB 717 lLysLeuTrpTrpGlnLeuGluLysGluProileAlaGlyAlaGluThrPheTrpValAs 737
QY 2001 CGGCGCGCGCAACCGCGAGACCAAGATCGCGCAAGCGCGCTACGTGACCGCGCGCGCGCG 2060
DB 737 pGlyAlaAlaAsnArgGluThrLysileGlyLysAlaGlyTrpValThrAspArgGlyAr 757
QY 2061 GCAGAAGATCGTGAGCTGACCGCGAGACCAACCAACAGAGACCGAGCTGCGAGCGCATCCA 2120
DB 757 gGlnLysileValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaileGl 777
QY 2121 GCTGCGCGCTGCGAGCAGCGCGCGAGGTGAACATCGTGACCGCGCGCGCGCGCGCGCGCT 2180
DB 777 nLeuAlaLeuGlnAspSerGluSerGluValAsnileValThrAspSerGlnTrpAlaLe 797
```

QY	2181	GGGATCATTCAGGCCAGCCGACCAAGAGCGAGCGAGCTGGTGAACCCAGATCATCGA	2240
Db	797	uGlyllelleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnlleleGI	817
QY	2241	GCAAGTATCAAGAGAGAGAGTGTACCTGAGCTGGTGGTCCCGCCCAACAGGGCATCGG	2300
Db	817	uGlnLeulleLysGluArgAlaTyrLeuSerTrpValProAlaHisLysGlylleGI	837
QY	2301	CGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGTGGCTG	2352
Db	837	yGlyAspGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu	854
RESULT 12			
AAO30963			
ID	AAO30963 standard; protein; 1003 AA.		
XX	AC	AAO30963;	
XX	DT	06-OCT-2003 (first entry)	
XX	DE	HIV pol protein.	
XX	XX	Human immunodeficiency virus; HIV; pol; IV9; immunotherapy; infection;	
KW	KW	epitope; vaccine; enzyme.	
XX	XX	Human immunodeficiency virus.	
OS	XX	WO2003051285-A2.	
PN	XX	26-JUN-2003.	
PD	XX	29-OCT-2002; 2002WO-US034620.	
PF	XX	29-OCT-2001; 2001US-0345116P.	
PR	XX	(GENZ ) GENZYME CORP.	
XX	PA	(GENO ) GEN HOSPITAL CORP DBA MASSACHUSETTS GEN.	
XX	XX	Nicolette CA, Walker BD;	
PI	XX	WPI; 2003-541603/51.	
DR	XX	Novel synthetic therapeutic peptide compounds useful for inducing an	
XX	PT	immune response in a mammal, specifically cross-react with HIV IV9	
PT	PT	epitopes.	
PS	XX	Claim 8; Page 65-67; 67pp; English.	
XX	XX	The invention relates to synthetic therapeutic peptide compounds that	
CC	CC	specifically cross-react with HIV (human immunodeficiency virus) pol	
CC	CC	protein IV9 epitopes. The invention is useful for treating HIV infection.	
CC	CC	The invention is also used in immunotherapy and as vaccines. The present	
CC	CC	sequence is HIV pol protein	
XX	SQ	Sequence 1003 AA;	
Alignment Scores:			
Pred. No.:	4.9e-199	Length:	1003
Score:	3535.00	Matches:	659
Percent Similarity:	96.12%	Conservative:	34
Best Local Similarity:	91.40%	Mismatches:	24
Query Match:	77.13%	Indels:	4
DB:	6	Gaps:	2
US-09-610-313B-30 (1-2469) x AAO30963 (1-1003)			
QY	220	TTCTTCGGCAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGCAGCGACG	279
Db	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
QY	280	AACCGCGCCAAACGCCCCACCGAGCTGCAGGTG-----CGCGCGCAACACCC	333

Db	21	ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer	40
QY	334	CGCAGCGAGCGCGCGCCGAGCGCAGGCGACCCCTG-----AACTTCCCGCAGATCAC	387
Db	41	ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr	60
QY	388	CTGTGCGCAGCGCCCTGGTGGAGCATCAAGGTGGCGGCAGATCAAGAGAGCCCTGCTG	447
Db	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu	80
QY	448	GACACCGCGCGCCGACGACACCGTGTGGAGAGATGAGCTGCCCGCAGAGTGAAGCC	507
Db	81	AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro	100
QY	508	AGATGATCGCGCGCATCGCGCTTCATCAAGTGGCGGCAGTACGACAGATCTCTGATC	567
Db	101	LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle	120
QY	568	GAGATCTGGCGCAAGAGCCATCGCGCCTGCTGATCGCGCCCGCCCGCTGAACATC	627
Db	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
QY	628	ATCGCGCGCAACATGCTGACCCAGCTGGGTGGTGCACCTGAATCTCCCATCAGCC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
QY	688	GAGACGTCGCCGTGAAGCTGAAGCCCGCATGAGCGCGCCCGCCCAAGGTGAAGAG	747
Db	161	GluThrValProValLysLeuLysProGlyMeAspGlyProLysValLysGlnTrpPro	180
QY	748	CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAAGAGGGC	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
QY	808	AGATCAACCAAGATCGCGCCCGGAGAACCTTACACACCCCGCTGTGTCATCAAGA	867
Db	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220
QY	868	AAGGACAGCACCAAGTGGCGAGCTGCTGCTCCGCGAGCTGAACAGCGCACCAG	927
Db	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
QY	928	GACTTCTGGAGGTGCGAGCTGGGCATCCCCACCCCGCGCTGAAGAGAAGAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
QY	988	GTGACCGTGTGACGTGGCGAGCGCTACTTACGGTGCCTTCCGCGAGGACTTCCGC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg	280
QY	1048	AGATACACCGCTTCCACCATCCCGCATCAACAGAGACCCCGCGCATCCGCTACGAC	1107
Db	281	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	340
QY	1228	GACCTGTGCTGGCGCAGCTGGAGATCGCGCAGCAGCCCGCGCAGAGATCGAGAGCTG	1287
Db	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
QY	1288	CGCAAGCACCTGTGCTGGCGCTTCCACCCCGCGCAAGAGACCAAGAGAGAGCC	1347
Db	361	ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro	380
QY	1348	CCCTTCTGTGATGGCTACGAGCTGCACCCCGCAGCAAGTGGACCGTGCAGCCCATCG	1407
Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpValGlnProIleVal	400



```
Qy 1408 CTGCCCGAGAGGAGCTGACCGTGAACCAATCCAGAGCTGGTGGCGCAAGCTGAAC 1467
Dy 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy 1468 TGGGCCAGCCAGATCACTACCGCGCATCAAGTGGCGCAGCTGTCAAGCTGTCTGGCGGC 1527
Dy 421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
Qy 1528 GCCAAGCCCTGACCGCATCTGTCCTCCCTGACCCAGAGGCGGAGCTGGAGCTGCCCGAG 1587
Dy 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
Qy 1588 AACCGCAGATCTCGCGCAGCCCGTGCACCGCTGTACTACGACCCCGCAGCAAGACTG 1647
Dy 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
Qy 1648 GTGCCCGAGATCCAGAGCAGGCGCACACAGTGGACCTACCAAGATCTACCAAGAGCCC 1707
Dy 481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
Qy 1708 TTCAAGAACCTGAAAGACCGCACTAGCCAAAGATGGCCACCGCCCAACCAACGAGCTG 1767
Dy 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
Qy 1768 AAGCAGCTGACCGAGCGCTGCAGAGATCCCATGGAGAGCATCTGTGATCTGGGCAAG 1827
Dy 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540
Qy 1828 ACCCCCAAGTCCCGCTGCCCATCCAGAGAGACCTGGGAGACCTGGTGACGACTAC 1887
Dy 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
Qy 1888 TGGCAGCGCCACTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTCAAGCTG 1947
Dy 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
Qy 1948 TGTACCACTGGAGAGGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGGCGCC 2007
Dy 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
Qy 2008 GCCAACCGCAGACCATGATCGGCAAGCGCGCTACGTACCGACCGCGGCGCGCAGAG 2067
Dy 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
Qy 2068 ATCGTGAGCTGACCGCAGACCAACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGCC 2127
Dy 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
Qy 2128 CTCGAGGACAGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2187
Dy 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy 2188 ATCAGCGCCAGCCGACAGAGCAGGAGCGAGCTGTGTGAACCATCATCGACGACTG 2247
Dy 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
Qy 2248 ATCAAGAGAGAGAGTGTACTCGAGCTGGTGCGCCGCCACAGAGGCATCGGCGGCAAC 2307
Dy 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
Qy 2308 GAGCAGATCGACAGCTGGTGAGCAAGGCGCATCCGCAAGGTGTCTTCTGGAGCGGCATC 2367
Dy 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
Qy 2368 GAT 2370
Dy 721 Asp 721
RESULT 13
ADN36413
ID ADN36413 standard; protein; 1003 AA.
XX
```

```
AC ADN36413;
XX 15-JUL-2004 (first entry)
DT XX HIV protein for anti-HIV vaccine.
DE XX anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX KW Human immunodeficiency virus 1.
OS XX WO2004035006-A2.
XX PN 29-APR-2004.
XX PD 17-OCT-2003; 2003WO-US033112.
XX PF 18-OCT-2002; 2002US-0419465P.
XX PR (AARO-) AARON DIAMOND AIDS RES CENT.
XX PA Huang Y, Ho DD, Chen Z;
XX PI WPI: 2004-348328/32.
XX DR N-PSDB; ADN36412.
XX CC Nucleic acid vector comprising at least one HIV sequence operably linked
PT to a promoter and encoding a protein that does not assemble into viral
PT particles, useful in immunizing a subject against HIV infection.
XX PS Disclosure; SEQ ID NO 27; 166pp; English.
XX CC The invention relates to a nucleic acid vector comprising at least one
CC HIV sequence operably linked to a promoter and encoding a protein that
CC does not assemble into viral particles. The nucleic acid vector is useful
CC in immunizing a subject against HIV infection. This sequence corresponds
CC to a nucleic acid used in the invention.
XX SQ Sequence 1003 AA;
Alignment Scores:
Pred. No.: 4,9e-199 Length: 1003
Score: 3535.00 Matches: 659
Percent Similarity: 96.12% Conservative: 34
Best Local Similarity: 91.40% Mismatches: 24
Query Match: 77.13% Indels: 4
DB: 8 Gaps: 2
US-09-610-313B-30 (1-2469) x ADN36413 (1-1003)
```

```
Qy 220 TTCTTCCGCGAGGACCTGGCTTCCCGCAGGCGCAAGCCCGCGAGTTCCCGAGCGAGCG 279
Dy 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy 280 AACCGCGCCCAACGCCCCACCGCGCGAGCTGCAGGTG-----CGCGGCGCAACCCC 333
Dy 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
Qy 334 CGCAGCGCAGCGCGCGCGCGAGCGCGCAGCGCACCTCGT-----AACTTCCCGCAGATCAC 387
Dy 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
Qy 388 CTGTGGCAGCGCCCGCTGGTGAGCATCAAGTGGCGCGCCAGATCAAGAGCGCCCTGTG 447
Dy 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCGCGCAGCACCGCTGCGAGAGATGAGCTGCCCGGCAAGTGGAGCGCC 507
Dy 81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100
Qy 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGTGGCGCGCAGTACGACCGACGATCTGTATC 567
Dy 101 LysMetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120
```

```
QY 568 GAGATCTGCGCAAGGACCATCGGCACCGTGTGATCGGCCCCACCCCGGTGAACATC 627
DB 121 GluIleCysGlyHisAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGGCGCAACATGTGACCCAGCTGGGTGACCCCTGAACCTTCCCATCAGCCCATC 687
DB 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACGCTGCCGTGAGCTGAAGCCCGGATCGGATCGGACCGCCCAAGTGAGCAGTGC 747
DB 161 GluThrValProValIleLeuLeuProGlyMetAspGlyProIleValIleGlyPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCGCTTCACCGCCATCTGCGAGGAGATGAGAAAGAGGC 807
DB 181 LeuThrGluGluIleLeuValGluIleCysThrGluMetGluLeuGly 200
QY 808 AAGATCACCAAGATCGGCGGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAG 867
DB 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGACTTCGCGAGCTGACAGCGCACCCAG 927
DB 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGGAGGTGACGTGGGCATCCCGCACCCCGCGCTGGAAGAAAGAAAGAGC 987
DB 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGCTGTGACGTGGCGAGCGCTACTTCAGCGTGGCCCTGACGAGGACTTCGCG 1047
DB 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGCAGCATCAACAGGAGACCCCGCATCCGCTACCCAG 1107
DB 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGTGCGCCAGGCTCGAAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACC 1167
DB 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCCAGTACATGGAC 1227
DB 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1228 GACTGTACGTGGCGAGCGACTGGAGATCGGCAGCAGCCCGCCAGATCGAGGAGCTG 1287
DB 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1288 CGCAAGCACCTGTGCGCTGGGCTTCACCAACCCCGCAGCAAGACCAAGAGAGAGGCC 1347
DB 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
QY 1348 CCTTCTGTGGATGGGCTACGAGTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAG 1407
DB 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1408 CTGCGCGAGAGAGAGCTGACCGTGAACAGCATCCAGAGCTGGTGGCAGCTGAAC 1467
DB 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1468 TGGGCGAGCAGATCTACCCCGGCGCATCAAGGTGGCGGAGCTGTGCAAGCTGTCGCGGC 1527
DB 421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1528 GCCAAGCCCTGACCGACATCGTGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAG 1587
DB 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1588 AACCGGAGATCTGCGGAGCCGCTGACGGGTGTACTACGACCCCGCAGCAGGACCTG 1647
DB 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1648 GTGGCGGAGATCCAGAAGCAGGGGCCACGACGAGTGGACCTTACCAGATCTACCGAGGCC 1707
```

```
DB 481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
QY 1708 TTCAAAGAACTGAAGACCGGCAAGTACGCCAAGATGGCAGCCGCCACACCAACGAGCTG 1767
DB 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
QY 1768 AAGCAGCTGACCGAGGCGCTGCGAAGAGATCGCCATCGAGAGCATCGTATCTGGGCAAG 1827
DB 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540
QY 1828 ACCCCCAAGTTCGCTGCCCATCCAGAAAGAGACCTGGGAGACCTGGTGAGCCGACTAC 1887
DB 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr 560
QY 1888 TGGCAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCTGTGTGAAGCTG 1947
DB 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580
QY 1948 TGGTACAGCTGGAGAAAGAGGCCATCATCGGCGCGGAGACCTTCTACGTGGACGGCGCC 2007
DB 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2008 GCCAACCGCGAGACCAAGATCGCAAGCGCGCTACGTGACCAACCGCGGCGCGCAGAG 2067
DB 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2068 ATCGTGAGCTGACCGAGACCCACCAACGAGACCGAGCTGCGAGGCCATCCAGCTGGCC 2127
DB 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
QY 2128 CTCGACGACAGCGCGCAGCGAGGTGAACCTGTCACCGACAGCAGCTACGCTGGGCATC 2187
DB 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2188 ATCCAGGCCAGCCCAAGACGAGAGCGAGCTGGTGAACCAAGATCATCGAGCAGCTG 2247
DB 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
QY 2248 ATCAAGAGGAGAGGTGTACCTGAGCTGGTGGTCCGCCACCAAGGGCATCGCGGCAAC 2307
DB 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700
QY 2308 GAGCAGATCGACAAGCTGGTGAAGAGGCGATCCGCAAGGTGCTGTCTCTGGACGGCATC 2367
DB 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2368 GAT 2370
DB 721 Asp 721
RESULT 14
AAP81861
ID AAP81861 standard; protein; 1002 AA.
XX
AC AAP81861;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 16-DEC-1990 (first entry)
XX
XX Sequence encoded by LAV MA L POL gene.
XX HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
XX Lymphadenopathy-associated virus.
XX WO8707906-A.
XX
XX 30-DEC-1987.
XX
XX 22-JUN-1987; 87WO-EP000326.
XX
```



Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
 QY 1831 CCAGTTCGGCTCCGTCATCCAGAGGACACCTGGGAGACCTGGGACGACTACTGG 1890  
 Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTrp 560  
 QY 1891 CAGGACACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGG 1950  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580  
 QY 1951 TACCAGCTGGAAGAGGCCATCATCGCGCCGAGACCTTCTACGTGGACGCGCGCC 2010  
 Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 QY 2011 AACCGGAGACCAAGATCGCAAGCGCGGTAGTGAACCGGCGGCGGCGGAGATC 2070  
 Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 QY 2071 GTGAGCTGACCGAGACCAACCAAGAACACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2130  
 Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
 QY 2131 CAGGACAGCGCGCAGCGAGTGAATCGTGAACCGACGACGACGACGCGCTGGGCATCATC 2190  
 Db 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 QY 2191 CAGGCCAGCGCCAGACGAGCGAGCGAGCTGGTGAACCGATCATCGAGCGCTGATC 2250  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680  
 QY 2251 AAGAAGGAGAGGTGTACCTGAGTGGTGGTCCCGCCACAAAGGCGCATCGCGCGCAACGAG 2310  
 Db 681 GlnLysAspLysValLysLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnGlu 700  
 QY 2311 CAGATCGACAAGCTGGTGTGACAAAGGCGCATCGCAAGGTGCTGTCTCTGGACGCGCATCGAT 2370  
 Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720  
 RESULT 15  
 ABR55489  
 ID ABR55489 standard; protein; 1003 AA.  
 XX  
 AC ABR55489;  
 XX  
 DT 11-AUG-2003 (first entry)  
 XX  
 DE Amino acid sequence of a HIV pol protein.  
 XX  
 KW Polymorphic marker; host response; HIV; AIDS related virus; vaccine;  
 KW hepatitis related virus; HCV; HBV; drug resistance; pol.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 PN WO2003035097-A1.  
 XX  
 PD 01-MAY-2003.  
 XX  
 PF 23-OCT-2002; 2002WO-AU001450.  
 XX  
 PR 23-OCT-2001; 2001AU-00008425.  
 XX  
 PA (EPIP-) EPIPOT PTY LTD.  
 XX  
 PI Mallal S;  
 XX  
 DR WPI; 2003-449231/42.  
 XX  
 PT Determining the influence of variation in host genes on the selection of  
 PT microorganisms with protein substitutions, comprises typing individuals  
 PT of a cohort infected with a microorganism for an intrinsic polymorphic  
 PT marker.  
 XX  
 PS Claim 22; Page 88-89; 157pp; English.

XX  
 CC The specification describes a method of determining the influence of  
 CC variation in host genes on selection of microorganisms with protein  
 CC substitutions. The method comprises typing all individuals of a  
 CC population of patients infected with a microorganism for at least one  
 CC selected intrinsic polymorphic marker involved in the host response to  
 CC the presence of the microorganism. The method is useful for examining  
 CC selective pressures confronting a wide range of organisms that exhibit  
 CC pathogenic traits in a host, such as bacteria, fungi, mycobacterium,  
 CC viruses and virus-like particles; for examining microorganisms that have  
 CC adapted to evolve rapidly, including HIV and AIDS related viruses and the  
 CC hepatitis related viruses such as HCV and HBV. The method is useful for  
 CC designing a vaccine to prevent or delay the emergence of drug resistance  
 CC in patients treated with a particular drug specific for a microorganism  
 CC where the drug affects the replication of the microorganism at the  
 CC nucleotide or amino acid level. The present sequence represents a HIV pol  
 CC protein, which is expected to provide optimal cytotoxic T lymphocyte  
 CC (CTL) induced therapeutic protection to the cohort examined in that study  
 XX  
 SQ Sequence 1003 AA;  
 Alignment Scores:  
 Pred. No.: 7,35e-199 Length: 1003  
 Score: 3532.00 Matches: 658  
 Percent Similarity: 96.26% Conservative: 36  
 Best Local Similarity: 91.26% Mismatches: 23  
 Query Match: 77.07% Indels: 4  
 DB: 6 Gaps: 2  
 US-09-610-313B-30 (1-2469) x ABR55489 (1-1003)  
 QY 220 TTCTTCGGGAGGACCTGGCTTCCCGGAGGCAAGCCGCGAGTTCCCGAGGAGCAG 279  
 Db 1 PhePheArgGlnAsnLeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGln 20  
 QY 280 AACCGGCGCAACAGCCCGACCGAGCTCGAGTGCAGCGCGGCGAC-----AACCCC 333  
 Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGluAspAsnAsnSer 40  
 QY 334 CGCAGCGAGCG 387  
 Db 41 ThrSerGlyAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
 QY 388 CTGTGCGAGCG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCG 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCG 567  
 Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleIleIle 120  
 QY 568 GAGATCTGCGGCAAGAGCG 627  
 Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGCGCGCAACATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687  
 Db 141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACGCTGCG 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGAGGAGAGATCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCACAAGATCG 867

201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
Db  
868 AAGGACACCAAGTGGCGAAGCTGTGACTTCCCGAGCTGAACAAGCGCACCCAG 927  
Qy  
221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
Db  
928 GACTCTCTGGAGGTGACGTGGCATCCCCACCCCGCGCGCTGAAGAAGAAGAGC 987  
Qy  
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
Db  
988 GTGACCGTCTGGAGCGTGGCGAGCCCTTACTTCCAGCGTGGCCCTGGAGAGACTTCCG 1047  
Qy  
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
Db  
1048 AAGTACACCGCTTACCATCCAGCATCAACAGAGACCCCGCGCATCCGCTTACCAG 1107  
Qy  
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
Db  
1108 TACACGTGTGCTCCCGAGGTGAAGGGCAGCCCGAGCATCTTCCAGAGCATGACC 1167  
Qy  
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
Db  
1168 AAGATCTGTGAGCCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACCACTACATGAC 1227  
Qy  
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
Db  
1228 GACTGTGCTGGGACGACCTGGAGATCGGCGAGCACCGCGCAAGATCGAGGAGCTG 1287  
Qy  
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
Db  
1288 CGAAGCACCTGTGCTGGGTGGTTCACACCCCGAGCAAGAGACCGAGAGAGCC 1347  
Qy  
361 ArgGlnHisLeuLeuLysTrpGlyPheThrProAspLysLysHisGlnLysGluPro 380  
Db  
1348 CCTTCTCTGTGATGGGTACGAGCTGCACCCCGAGTGGAGCGGTGCGCCCATCGAG 1407  
Qy  
381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
Db  
1408 CTGCCCCAGAGAGAGCTGGACCGTGAACACATCCAGAGCTGTGGGCAAGCTGAAC 1467  
Qy  
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
Db  
1468 TGGCCAGCCAGATCTACCCCGGATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGGC 1527  
Qy  
421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
Db  
1528 GCCAAGCCCTGACCGACATGTCCTCCCTGACCGAGGCGCGAGCTGGAGCTGGCCGAG 1587  
Qy  
441 ThrLysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGlu 460  
Db  
1588 AACCGCAGATCTCTGCGCAGCGCTGACGCGCTGTACTACGACCCCGAGAGGACCTG 1647  
Qy  
461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480  
Db  
1648 GTGCCGAGATCCAGAGAGGCGCCAGCAGTGGACCTTACAGATCTTACAGAGGCC 1707  
Qy  
481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500  
Db  
1708 TTCAAGAACCTGAGACCGCGNAGTACGCCAAGTGGCCACCGCCCAACACAGAGCTG 1767  
Qy  
501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
Db  
1768 AAGCAGCTGACCGAGGCGGTGCAAGATCGCATGGAGAGCATCGTGATCTGGGGCAAG 1827  
Qy  
521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
Db  
1828 ACCCCCAAGTTCGCGCTGCCATCCAGAGAGACCTTGGAGACCTTGGTGGACCGACTAC 1887  
Qy  
541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560  
Db  
1888 TGGCAGGCCACTGGATCCCGAGTGGAGTGTGTGAACACCCCGCCCTGGTGAAGCTG 1947  
Qy  
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
Db

1948 TGGTACCAGCTGGAGAAAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCC 2007  
Qy  
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
Db  
2008 GCCAACCCGAGAGACCAAGATCGGCAAGCCCGCTACGTGACCGACCGGGCGCGCAGAAG 2067  
Qy  
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620  
Db  
2068 ATCTGTAGCTGACCGAGACCAACCAAGAGACCGAGCTGACGGCCATCCAGCTGGCC 2127  
Qy  
621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
Db  
2128 CTGACGACACCGCGAGCGAGGTCAACATCTGTGACCGACACCGACCTTGGGCATC 2187  
Qy  
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
Db  
2188 ATCCAGGCCCGCGAGAGCGAGCGAGCGAGCTGTGAACACAGATCATCGAGCAGCTG 2247  
Qy  
661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
Db  
2248 ATCAAGAGGAGAGGTGTACTGAGCTGGGTGCGCCGCCCAAGGGCATCGCGGCAAC 2307  
Qy  
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
Db  
2308 GAGCAGATCCAGACAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTGTCTTCTGGACGGCATC 2367  
Qy  
701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
Db  
2368 GAT 2370  
Qy  
721 Asp 721

Search completed: June 2, 2005, 04:41:43  
Job time : 256.242 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 11:33:41 ; Search time 167.833 Seconds  
(without alignments)

11351.628 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 4573

Sequence: 1 gtcagccaccatgcccga.....gggctagaccggtgaattc 2463

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09610313/runat\_31052005\_151136\_151139/app\_query.fasta\_1.7893  
-DB-A Geneseq 16Dec04 -QPMF=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USR=US09610313 @CGN 1 1 557 @runat 31052005 151136 151139 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A Geneseq 16Dec04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3766.5	82.4	1435	8 ADS19486	AdS19486 Modified
2	3744.5	81.9	3193	8 ADP84803	Adp84803 HIV-1 hyb
3	3711.5	81.2	1457	8 ADN36406	Adn36406 Human pro
4	3620	79.2	1003	3 AAB69289	Aab69289 HIV-1 non
5	3611	79.0	999	3 AAB69286	Aab69286 HIV-1 non
6	3606	78.9	1005	3 AAB69287	Aab69287 HIV-1 non
7	3570	78.1	998	6 AAE37601	Aae37601 HIV-1 sub
8	3563	77.9	3025	4 AAB86169	Aab86169 HIV-1 sub
9	3521	77.0	854	5 AAM48949	Aam48949 HIV-1 sub
10	3517.5	76.9	1000	3 AAB69282	Aab69282 HIV-1 non

11	3499.5	76.5	1002	2 AAW72993	Aaw72993 HIV isola
12	3497	76.5	1003	6 AAO30963	Aao30963 HIV pol p
13	3497	76.5	1003	8 ADN36413	Adn36413 HIV prote
14	3494.5	76.4	1002	1 AAP81861	Aap81861 Sequence
15	3494	76.4	1003	6 ABR55489	Abr55489 Amino aci
16	3492	76.4	1003	1 AAP60420	Aap60420 Sequence
17	3492	76.4	1003	1 AAP70861	Aap70861 Sequence
18	3486	76.2	1003	3 AAY70602	Aay70602 Codon opt
19	3486	76.2	1003	3 AAY70601	Aay70601 Corrected
20	3484	76.2	1003	5 AAO19387	Aao19387 Lymphaden
21	3478.5	76.1	1003	2 AAR08060	Aar08060 HIV-1 pol
22	3475	76.0	1015	1 AAP60347	Aap60347 HTLV-III
23	3475	76.0	1015	2 AAR43867	Aar43867 HTLV-III
24	3475	76.0	1015	4 AAB85993	Aab85993 Amino aci
25	3472	75.9	1003	3 AAB10047	Aab10047 HIV-1 pol
26	3472	75.9	1003	3 AAY70600	Aay70600 Wild type
27	3471	75.9	1016	2 AAR08062	Aar08062 AcNPV-HIV
28	3468	75.8	1015	2 AAR43875	Aar43875 HTLV-III
29	3464	75.7	1016	2 AAR08063	Aar08063 HIV-1 pol
30	3461	75.7	2033	2 AAR08056	Aar08056 HIV-1 pol
31	3457	75.6	1016	2 AAR08054	Aar08054 HIV-1 pol
32	3455	75.6	1003	3 AAB69284	Aab69284 HIV-1 non
33	3454	75.5	2033	2 AAR08055	Aar08055 HIV-1 pol
34	3453	75.5	1003	8 ADP20076	Adp20076 Human imm
35	3449	75.4	739	5 AAU11874	Aau11874 HIV pol p
36	3449	75.4	1003	1 AAP61508	Aap61508 Sequence
37	3449	75.4	1003	2 AAR29705	Aar29705 pol gene
38	3449	75.4	1003	3 AAY77301	Aay77301 HIV-1 (AT
39	3449	75.4	1003	5 AAE35789	Aae35789 ARV-2 (9B
40	3449	75.4	1003	6 ABU63185	Abu63185 Pol prote
41	3449	75.4	1491	1 AAP91048	Aap91048 Transcrip
42	3446	75.4	1003	3 AAB69279	Aab69279 HIV-1 non
43	3444	75.3	1003	6 ABU57552	Abu57552 AIDS asso
44	3442	75.3	1012	6 ABU63325	Abu63325 Human lym
45	3439.5	75.2	1003	2 AAR08059	Aar08059 HIV-1 pol

ALIGNMENTS

RESULT 1

ADS19486  
ID ADS19486 standard; protein; 1435 AA.

AC ADS19486;

DT 18-NOV-2004 (first entry)

DE Modified HIV-1 GagPol protein.

KW Gene delivery; epithelial cell; respiratory system; mammal;  
KW lentiviral packaging system; expression vector; gagpol; gene expression;  
KW lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;  
KW cystic fibrosis; CF; deltaPD; transepithelial potential difference;  
KW gene therapy; HIV-1; mutant; mutain.

OS Human immunodeficiency virus 1.

OS Synthetic.

PN US2004037780-A1.

PD 26-FEB-2004.

XX 23-AUG-2002; 2002US-00226638.

XX 19-NOV-2001; 2001AU-00008942.

XX (PARS/) PARSONS D.

XX (ANSO/) ANSON D.

XX (LIMB/) LIMBERIS M.

XX (FULL/) FULLER M.

XX Parsons D, Anson D, Limberis M, Fuller M;



DR WPI; 2004-191004/18.  
 DR N-PSDB; ADS19485.  
 XX Delivering exogenous (e.g. CFTR) genes to epithelial cells in the  
 PT respiratory tract using a penetration agent and a lentiviral packaging  
 PT system, useful for treating e.g. cystic fibrosis.  
 XX  
 PS Disclosure; Fig 16; 45pp; English.  
 XX  
 CC The present invention relates to a method of delivering one or more  
 CC exogenous genes for expression in an epithelial cell in the respiratory  
 CC system of a mammal to give persistent expression of the gene in the  
 CC epithelial cell, using a lentiviral packaging (expression vector) system.  
 CC Also disclosed are (i) a recombinant lentiviral packaging system,  
 CC comprising a first nucleic acid molecule including a gag gene sequence  
 CC encoding a gag protein, and a second nucleic acid molecule including a  
 CC gagpol gene sequence encoding a GagPol protein (the gagpol gene sequence  
 CC has degenerative nucleotide changes in the frame shift sequence AUUUUU  
 CC to reduce the chance of a frame shift which switches expression of the  
 CC Gagpol protein to the gag protein in wild type lentivirus and the  
 CC packaging system additionally comprises a lentiviral vector nucleic acid  
 CC molecule which does not encode either the gag gene or the gagpol gene or  
 CC both, and (ii) a recombinant nucleic acid molecule encoding a lentiviral  
 CC gagpol gene having degenerative nucleotide changes in the frame shift  
 CC sequence AUUUUU to reduce the chance of a frame shift which switches  
 CC expression of the Gagpol protein to the gag protein in the wild type  
 CC lentivirus. The exogenous gene is expressed in sufficient numbers of  
 CC cells and amounts to provide an ameliorating effect for a respiratory  
 CC condition. The exogenous gene is cystic fibrosis transmembrane  
 CC conductance regulator (CFTR) and the condition is cystic fibrosis  
 CC ((CF) the expression of the CFTR gene is sufficient to provide a  
 CC significant shift of a reduced deltaPD (transepithelial potential  
 CC difference) back to normal levels in the mammal). The cell is non-  
 CC terminally differentiated and is capable of differentiating into 2 or  
 CC more cell classes e.g. ciliated cells, non-ciliated cells, secretory  
 CC cells and basal cells. The exogenous gene may be enzymic. The gagpol gene  
 CC sequence has additionally degenerate nucleotide substitutions which  
 CC destabilise the hairpin structure associated with the frame shift event.  
 CC The gag gene is a truncation of the wild type gagpol gene so that it can  
 CC no longer be translated to form Gagpol. The lentivirus is HIV or HIV  
 CC derived. The method and recombinant lentiviral packaging system may be  
 CC used for delivering nucleic acids to epithelial cells in the respiratory  
 CC system of a mammal. In particular they may be used to deliver the CFTR  
 CC gene for the treatment of CF. It has been found that the delivery of a  
 CC recombinant lentivirus carrying an exogenous gene to the respiratory  
 CC system, following the delivery of a non-toxic amount of a penetration  
 CC enhancer can provide persistence of expression of a gene product. The  
 CC present sequence represents modified HIV-1 Gagpol protein.  
 XX  
 SQ Sequence 1435 AA;

Alignment Scores:  
 Pred. No.: 2, 87e-209 Length: 1435  
 Score: 3766.50 Matches: 702  
 Percent Similarity: 94.31% Conservative: 44  
 Best Local Similarity: 88.75% Mismatches: 38  
 Query Match: 82.36% Indels: 7  
 DB: 8 Gaps: 4

US-09-610-313B-31 (1-2463) x ADS19486 (1-1435)

QY	13	ATGCGCGAGCCATGAGCCAGGCCACC--AGCGCAACATCTGATGCGAGCGCAGCAAC	69
Db	363	LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn	382
QY	70	TTCAAGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGGAGGGCCCATCGCC	129
Db	383	PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla	402
QY	130	CGCAATGCGCGCCCCCGCAGAGAGGGCTGCTGGAGTGGCGGCAAGGAGGGCCACCG	189
Db	403	LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln	422
QY	190	ATGAGGAGTCTCACCCGAGCGCCAGGCAACTTCTTCCGGAGGACCTGGCTTCCCCCAG	249
Db	423	MetLysAspCysThrGluArgGlnAlaAsnPheLeuArgGluAspLeuAlaPheProGln	442
QY	250	GGCAAGCCCGGAGTTCCCGAGCAGAGAACCCGCGCAACAGCCCCACCGAGCCGAG	309
Db	443	GlyLysAlaArgLysPheSerGluGlnThrArgAlaAsnSerProIleArgArgGlu	462
QY	310	CTGCAGGTG-----CGCGGCACACCCCGAGCGAGCGCGCGCCGAGCGCCAGGGC	363
Db	463	ArgGlnValTrpArgArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGly	482
QY	364	ACCTG-----AACTTCCCCAGATCACCTGTGCGAGCGCCCTGGTGGACATCAAG	417
Db	483	ThrValSerPheSerPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLys	502
QY	418	GTGGCGGCGCAGATCAAGGAGGCCCTGTGGACACCGCGCGCCGACACCGTGTGGAG	477
Db	503	IleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGlu	522
QY	478	GAGTACGCTGCCCGGCAAGTGGAGGCCACAGATCATCGCGGCGCATCGCGGCTTCATC	537
Db	523	GluMetAsnLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGlyPheIle	542
QY	538	AAGTGCAGCATACACACAGATCTCTGATCGAGATCTGCGGCAAGAGGCCATCGGCACC	597
Db	543	LysValArgGlnTrpAspGlnIleProIleGluIleCysGlyHisLysAlaIleGlyThr	562
QY	598	GTGCTGATCGGCCCAACCCCGTGAACATCATCGCGCGCCGACACCGTGTGGAG	657
Db	563	ValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGly	582
QY	658	TGCACCTGAACTTCCCATCAGCCCATCGAGAGCGTGCCTGAGCTGAGAGCCCGGC	717
Db	583	CysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGly	602
QY	718	ATGACGCGGCCCAAGTGAAGCAGTGGCCCTCAGCCAGAGAGATCAAGGCCCTGACC	777
Db	603	MetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuVal	622
QY	778	GCATCTCGAGAGATGAGAGAGGAGGAGATCACCAGATCGGCGCCCGAGAACCC	837
Db	623	GluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnPro	642
QY	838	TACAAACCCCGTGTCCCATCAAGAGAGAGCAGCAGCAGGAGGCGCAAGTGGCGCAAGCTGGT	897
Db	643	TyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuVal	662
QY	898	GACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCGAGTGGCGCATCCCC	957
Db	663	AspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIlePro	682
QY	958	CACCCCGCGGCTGAAGAGAGAGCGTGCCTGAGTGCCTGAGCGCGAGCGCTTAC	1017
Db	683	HisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyr	702
QY	1018	TTTCAGCGTCCCTGCGAGCGGACTTCCGCAAGTACACCGCTTCCATCCCGCAGCATC	1077
Db	703	PheSerValProLeuHisGluAspPheArgLysTrpThrAlaPheThrIleProSerIle	722
QY	1078	AACAAAGAGACCCCGGCATCCCTACCAAGTACAAGTGTGCTGCCCGAGGGCTGGAAGGC	1137
Db	723	AsnAsnGluThrProGlyThrArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGly	742
QY	1138	AGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCTTGAGGCCCTTCGCGCGCCGCAAC	1197
Db	743	SerProAlaIlePheGlnSerSerMetThrThrIleLeuGluProPheArgLysGlnAsn	762
QY	1198	CCCGAGATCGTGTATCTTACCAG-----GCCCTCTGTAGTGGCGCAGCGACCTGGAGATC	1251
Db	763	ProAspLeuIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIle	782
QY	1252	GGCCAGCAGCCGCGCCAAGATCGAGGAGCTGCGCAAGACCTGTGCTGCGCTGGGCTTCACC	1311



QY 13 ATGGCCGAGCCATGACCGCCAGCCACGAGC---GCCAATCTCTGATGACGCCGACCAAC 69  
Db 363 LeuAlaGluAlaMetSerGlnValThrAsnProAlaThrIleMetIleGlnLysGlyAsn 382  
QY 70 TTCAAGGGCCCAAGCCATCATCAAGTGTCTCAACTGCGGCAAGAGAGGCCACATCGCC 129  
Db 383 PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysGluGlyHisIleAla 402  
QY 130 CGCAACTGCGCGCCCGCCGAGAGAGGGTGTCTGGAAGTGCAGGAGAGGGCCACAG 189  
Db 403 LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln 422  
QY 190 ATCAAGGACTGCACCGAGCCAGGCCAAC----- 219  
Db 423 MetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrpProSerHisLys 442  
QY 220 -----TTCTTCCCGCAGGAC 234  
Db 443 GlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPhePheArgGluAsp 462  
QY 235 CTGGCTTCCCGAGGCAAGGCCGAGTTCCTCCAGCAGCAGAACCGCGCCAAACAGC 294  
Db 463 LeuAlaPheProGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSer 482  
QY 294 ----- 294  
Db 483 ProProGluGluSerPheArgPheGlyGluGluThrThrThrProSerGlnLysGlnGlu 502  
QY 295 -----CCACACGAGCCGAGCTGCAGGTG----- 318  
Db 503 ProIleAspLysGluLeuTyrProLeuAlaProThrArgArgGluLeuGlnValTrpGly 522  
QY 319 CGCGCGGACAAACCCCGCAGCAGCGCGCGCGCAGCGCCAGCGCACCTGTAAC----- 372  
Db 523 ArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSer 542  
QY 373 -----TTCCCCCAGATCACCTGTGG 393  
Db 543 SerLeuArgSerLeuPheGlySerAspProSerSerGlnPheProGlnIleThrLeuTrp 562  
QY 394 CAGCGCCCTGTGTGATCAAGTGTGGCGGCAGATCAAGAGAGCCCTGTGTGGACACC 453  
Db 563 GlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeuAspThr 582  
QY 454 GGCGCCGACACACCTGTGTGGAGGATGAGCTGCCCGCGCAAGTGGAAAGCCCAAGATG 513  
Db 583 GlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysProLysMet 602  
QY 514 ATCGCGGCATCGCGCTTCATCAAGTGTGGCCAGTACGACAGATCCTGTGAGATC 573  
Db 603 IleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIle 622  
QY 574 TCGCGCAAGAGCCATCGGCACCGTGTGATCGGCCACCCCGCCGTAACATCATCGGC 633  
Db 623 CysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGly 642  
QY 634 CGCAACATGTCACCCAGCTGGCTGCACCTGAACTTCCCATCAGCCCCCATCGAGACC 693  
Db 643 ArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThr 662  
QY 694 GTCCCGTGAAGTGAAGCCCGCATGGACGGCCCGCCAGTGAAGCAGTGGCCCTGACC 753  
Db 663 ValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThr 682  
QY 754 GAGGAGAAGATCAAGGCCCTGACCGCCATCTCGAGGAGATGAGAGAGAGGCGCAAGATC 813  
Db 683 GluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGlyLysIle 702  
QY 814 ACCAAGTCGGCCCGAGAACCCCTTACAACACCCCGTGTTCGCCATCAAGAGAGAGGAC 873  
Db 703 SerLysIleGlyProGluAsnProLysAsnThrProValPheAlaIleLysLysLysAsp 722

QY 874 AGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCACCCAGGACTTC 933  
Db 723 SerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPhe 742  
QY 934 TGGAGGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAAGAGCGTGCACC 993  
Db 743 TrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysGlnLysLysSerValThr 762  
QY 994 GTCTCTGAGCGTGGCGACCGCTACTTCAGGTGCTCCCTGGACGAGGACTTCGCCAAGTAC 1053  
Db 763 ValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLysTyr 782  
QY 1054 ACCGCTTCAACATCCCCAGCACAACACAGACACCCCGGCATCCGCTACCACTAGTCAAC 1113  
Db 783 ThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsn 802  
QY 1114 GTCTCTCCCGAGGCTGGAAGGCGACCCAGCATCTTCAGACGACGATGACCAAGATC 1173  
Db 803 ValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnCysSerMetThrLysIle 822  
QY 1174 CTGGAGCCTTCCGCGCCCGCAACCCCGAGATCGTGTATCTACCAG-----GCCCCCTG 1227  
Db 823 LeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeu 842  
QY 1228 TAGCTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCAAGATCGAGAGCTGCGCAAG 1287  
Db 843 TyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuLeuArgGln 862  
QY 1288 CACTCTCTCGCTGGGCTTCCACCCCGCACAGAAGCACCAAGAGAGCCCTCTTC 1347  
Db 863 HisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPhe 882  
QY 1348 CTGTGATGGCTACAGCTGACCCCGACAGTGCACCGTGCAGCCCATCGAGCTGCC 1407  
Db 883 LeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleValLeuPro 902  
QY 1408 GAGAAAGAGAGCTGGACCGTGAACACATCCAAAGCTGGTGGCAAGCTGAACCTGGGCC 1467  
Db 903 GluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAla 922  
QY 1468 AGCAGATCTACCCCGCATCAAGTGTGCGCCAGCTGTGAAGTGTGCGCGCGCAAG 1527  
Db 923 SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLys 942  
QY 1528 GCCTGACCGACATCTGCGCCCTGACCGAGAGCGCGAGCTGCGCGAGAACCCG 1587  
Db 943 AlaLeuThrGluValValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArg 962  
QY 1588 GAGATCTCGCGAGCCCGTGCACCGCTGTACTACGACCCCGACAGGACCTGGTGGCC 1647  
Db 963 GluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIleAla 982  
QY 1648 GAGATCCAGAAGCAGGCGCCACCGACCTACAGATCTTACAGAGCCCTTCAAG 1707  
Db 983 GluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluProPheLys 1002  
QY 1708 AACCTGAAGACCGGCAAGTACGCAAGATGCGCACCGCCCGCACACCGAGCTGAAGCAG 1767  
Db 1003 AsnLeuLysThrGlyLysTyrAlaArgMetLysGlyAlaHisThrAsnAspValLysGln 1022  
QY 1768 CTGACCGAGCGCTGCAGAAGATCGCCATGAGAGATCGTGTATCTGGGCGAAGACCCC 1827  
Db 1023 LeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThrPro 1042  
QY 1828 AAGTCTCCGCTGCCATCCAGAGAGACCTGGGAGACCTGGTGGACCGCACTACTGGCAG 1887  
Db 1043 LysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGln 1062  
QY 1888 GCCACCTGTGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTGGAGCTGGTAC 1947  
Db 1063 AlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuLysTrpTyr 1082  
QY 1948 CAGCTGGAGAAGGAGGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGCCCAAC 2007

```

Db 1083 GlnLeuGluLysGluProIleGlyAlaGluThrPheTyrValAspGlyAlaAlaAsn 1102
Qy 2008 CGCGAGACCAAGATCGGCAAGCGCGCTACCTGACCGACCGCGCGCGCGAGAGATCGTG 2067
Db 1103 ArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValVal 1122
Qy 2068 AGCTGACCGAGACCAACCAAGACAGACCGAGCTGAGCGATCCAGTGGCCCTGAG 2127
Db 1123 ProLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142
Qy 2128 GACAGCGCGAGGAGTGAACATCGTGACCGACAGCAGCAGTACGCGCTGGGATCATCCAG 2187
Db 1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGln 1162
Qy 2188 GCCCAGCGCGAGAGCGAGCGAGCTGGTGTGAACACAGATCATCGACGCTGATCAAG 2247
Db 1163 AlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuLys 1182
Qy 2248 AAGGAGAAGGTGTACCTGAGCTGGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 2307
Db 1183 LysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGln 1202
Qy 2308 ATCCGACAGCTGTGAGCAAGGCGATCCGCAAGGTGCTGTTCTGGAGCGGATCGAT 2364
Db 1203 ValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1221

RESULT 3
ADN36406
ID ADN36406 standard; protein, 1457 AA.
AC ADN36406;
XX
DT 15-JUL-2004 (first entry)
XX
DE Human protein for anti-HIV vaccine.
XX
KW anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX
OS Homo sapiens.
XX
PN W02004035006-A2.
XX
PD 29-APR-2004.
XX
PF 17-OCT-2003; 2003WO-US033112.
XX
PR 18-OCT-2002; 2002US-0419465P.
XX
PA (AARO-) AARON DIAMOND AIDS RES CENT.
XX
PI Huang Y, Ho DD, Chen Z;
XX
DR WPI; 2004-348328/32.
XX
DR N-PSDB; ADN36405.
XX
CC Nucleic acid vector comprising at least one HIV sequence operably linked
PT to a promoter and encoding a protein that does not assemble into viral
PT particles, useful in immunizing a subject against HIV infection.
XX
PS Disclosure; SEQ ID NO 20; 166pp; English.
XX
CC The invention relates to a nucleic acid vector comprising at least one
CC HIV sequence operably linked to a promoter and encoding a protein that
CC does not assemble into viral particles. The nucleic acid vector is useful
CC in immunizing a subject against HIV infection. This sequence corresponds
CC to a peptide used in the invention.
XX
SQ Sequence 1457 AA;

Alignment Scores:
Pred. No.: 4.34e-206 Length: 1457
Score: 3711.50 Matches: 706

```

```

Percent Similarity: 91.86% Conservative: 28
Best Local Similarity: 88.36% Mismatches: 45
Query Match: 81.16% Indels: 21
DB: 8 Gaps: 4
US-09-610-313B-31 (1-2463) x ADN36406 (1-1457)

Qy 13 ATGCGCGAGCGCATGAGCCAGCGCACAGCGCCCAACATCTGTATGTCAGCGCGCAACCTTC 72
Db 383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnBhe 401
Qy 73 AAGGGCCCCAAGCGCATCATCAAGTGTTCACGTGCGGCAAGGAGGGGCCACATCCCGCGC 132
Db 402 LysGlySerLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 421
Qy 133 AACTGCGCGCGCCCCCGCAAGAGGGTGTCTGGAAGTCTCGGCAAGAGGGGCCACCATGATG 192
Db 422 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleAlaMet 441
Qy 193 AAGGACTGCACCGAGCGCGAGCCCAACTTCTTCGCGAGAGACCTGGCTTCCCCCAGCGGC 252
Db 442 LysAspCysThrGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProProHisLysG1 461
Qy 253 AAGGCGCGCGAGTTCCTCCAGCGAGCAGAACCGCGCCCAACAGCCCGCGCGCGAGCTG 312
Db 461 YArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProAlaGluSerPh 481
Qy 313 CAGGTGCGCGCGAGCAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 481 eGlyPheGluGluThrThrProAlaProLysGlnGluProLysAspArgGluProLeuTh 501
Qy 373 TTC-----CCCCAGATCACCCCTGTGTGCA 395
Db 501 rSerLeuLysSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpG1 521
Qy 396 CGCGCGCGCGTGTGAGCATCAAGTGGCGCGCGCGCATCAAGAGGCGCGCGCGCGCGCGCG 455
Db 521 nArgProLeuValSerIleArgValGlyGlyGlnIleLysGluAlaLeuLeu----- 538
Qy 456 CGCGCGAGCACACCGTGTGAGAGAGATGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515
Db 539 ----AspAspThrValLeuGluValAsnLeuProGlyLysTrpLysProLysMetI1 557
Qy 516 CGCGCGCATCGCGCTTCATCAAGGTGCGCGCATGACAGCATCTGTATGTCAGATCTG 575
Db 557 eGlyGlyIleGlyPheIleLysValArgLysTrpAspGlnIleProIleGluIleCy 577
Qy 576 CGCGAAGAGCGCATCGCGCGCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 635
Db 577 sGlyLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597
Qy 636 CAACATGCTGACCGCGTGGGCTGCACCTGAACTTCCCATCATGAGCGCGCGCGCGCGCGCG 695
Db 597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrI1 617
Qy 696 GCCCGTGAAGTGAAGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
Db 617 eProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrG1 637
Qy 756 GGAGAAGATCAAGCGCGTGGCGCGCATCTCGAGAGATGAGAGAGGCGCGCGCGCGCGCGCG 815
Db 637 uGluLysIleLysAlaLeuThrAlaIleCysAspGluMetGluLysGluGlyLysIleTh 657
Qy 816 CAAGATCGCGCGCGCGAGAACCCCTCAACACCGCGCGTGTTCGCCATCAAGAGAAGAGCAG 875
Db 657 rLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSe 677
Qy 876 CACCAAGTGGCGCAAGCTGTGGACTTTCGCGAGCTGAACAGCGCGCGCGCGCGCGCGCTCT 935
Db 677 rThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTr 697
Qy 936 GGAGGTGCGAGTGGCGCATCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995

```

Db 697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVa 717  
 QY 996 GCTGGAGTGGGAGCCCTACTTCCAGCGTCCCTCGAGAGAGACTTCCGCAAGTACAC 1055  
 Db 717 lLeuAspValGlyAspAlaTyrPheSerValProLeuTyrGluAspPheArgLysTyrTh 737  
 QY 1056 CGCTTTCCACCATCCAGCATCAACACAGAGACCCCGGCATCCGCTTACCAGTACCAACGT 1115  
 Db 737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVa 757  
 QY 1116 GCTGCCAGGCTGGAAGCGCAGCCCGACATCTTCCAGAGCAGCATGACCAAGATCCT 1175  
 Db 757 lLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777  
 QY 1176 GGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGTACTACCAG-----GCCCCCTGTA 1229  
 Db 777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyrGlyAspAspLeuTy 797  
 QY 1230 CTTGGGCGAGGACTGTGAGATCGGCAGACCCCGCCAGATCGAGGAGCTGCGCAAGCA 1289  
 Db 797 rValGlySerAspPheGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHi 817  
 QY 1290 CTTGCTGCGTGGGCTTACCACCCCGCAAGAACGACCAAGAGAGCCCTTCTTCT 1349  
 Db 817 sLeuLeuLysTyrGlyPheThrThrProAspLysLysHisGlnLysGluProPheLe 837  
 QY 1350 GTGGATGGGTACGAGCTGCACCCCGACAGTGGAGCGTGCAGCCCATCGAGCTGCCGA 1409  
 Db 837 utrMetGlyTyrGluLeuHisProAspLysTyrThrValGlnProIleGlnLeuProGl 857  
 QY 1410 GAAGGAGAGCTGGACCGTGAACACATCCAGAGCTGGTGGGCAAGCTGAACCTGGGCGCAG 1469  
 Db 857 uLysAspSerTyrThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877  
 QY 1470 CCAGATCTACCCCGCATCAAGTGGCCAGCTGTGTCAAGCTGCTGCGCGCCCAAGGC 1529  
 Db 877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897  
 QY 1530 CTTGACCGCATGCTGCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGGAGACCGCGA 1589  
 Db 897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGl 917  
 QY 1590 GATCTCGCGGAGCGCGTGCACCGCGTGTACTACGACCCCGCAGAGGACCTGTGTGGCCGA 1649  
 Db 917 uIleLeuLysGluProValHisGlyAlaTyrTyrAspProSerLysAspLeuIleAlaGl 937  
 QY 1650 GATCCAGACAGCGGCGCACAGCAGTGGACCTTACCATGATCTACAGAGCGCTTCAAGAA 1709  
 Db 937 uIleGlnLysGlnGlyAspGlnTyrThrTyrGlnIleTyrGlnGluProPheLysAs 957  
 QY 1710 CTTGAGACCGGCAAGTACCGCAAGATCGCACCGGCCACACCAACGACGTGAGCAGCT 1769  
 Db 957 nLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977  
 QY 1770 GACCGAGGCGGTGCAGAAATCCGCAAGATCGTGTGAGAGCATCTGTGGGCAAGACCCCA 1829  
 Db 977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTyrGlyLysIleProLy 997  
 QY 1830 GTTCCGCTCCCATCCAGAGAGACCTCGGAGACCTGTGAGACCGACTACTGGCAGGC 1889  
 Db 997 sPheArgLeuProIleProLysGluThrTyrGluThrArgTyrThrAlaTyrTrpGlnAl 1017  
 QY 1890 CACTGGATCCCGAGTGGAGTTCGTGACACCCCCCTCGTGAAGCTGTGTACCA 1949  
 Db 1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGl 1037  
 QY 1950 GCTGGAGAGAGCGCCCATCATCGCGCCGAGACTTCTACGTGAGCGCGCCGCAACCG 2009  
 Db 1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyrValAspGlyAlaAlaAsnAr 1057  
 QY 2010 CGAGACCAAGATCGGCAAGCGCGGTACTGTCAGCGACCGCGGCGCGGAGAGTCTGTGAG 2069  
 Db 1057 gGluThrLysMetGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSe 1077

QY 2070 CCTGACCGAGACCACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGGA 2129  
 Db 1077 rLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleCysLeuAlaLeuGlnAs 1097  
 QY 2130 CAGCGCGCAGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGC 2189  
 Db 1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAl 1117  
 QY 2190 CAGCGCCGACAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTGATCAAGAA 2249  
 Db 1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLy 1137  
 QY 2250 GGAGAGGTGTACTCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCAACCGACGACAT 2309  
 Db 1137 sGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyCysGlnGluGlnVa 1157  
 QY 2310 CGACAGCTGGTGACGACGAGCATCGCAGAGGTGCTGTCTCTCGAGCGGCATCGAT 2364  
 Db 1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175  
 RESULT 4  
 AAB69289  
 ID AAB69289 standard; protein; 1003 AA.  
 XX AAB69289;  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 DE HIV-1 non-subtype B clone 94IN476-104 pol protein.  
 XX  
 KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 XX  
 OS Human immunodeficiency virus 1.  
 XX  
 PN WC200026416-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 25-OCT-1999; 99WO-US024837.  
 XX  
 PR 02-NOV-1998; 98US-00184418.  
 XX  
 PA (UABR-) UAB RES FOUND.  
 XX  
 PI Hahn BH, Shaw GM, Gao F;  
 XX  
 DR WPI; 2000-365651/31.  
 XX  
 XX Claim 41; Fig 15; 131pp; English.  
 CC The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1003 AA;  
 Alignment Scores:  
 Pred. No.: 7,76e-201 Length: 1003  
 Score: 3620.00 Matches: 677  
 Percent Similarity: 98.05% Conservative: 26  
 Best Local Similarity: 94.42% Mismatches: 12

Query Match: 79.16% Indels: 2  
DB: 3 Gaps: 1  
US-09-610-313B-31 (1-2463) x AAB69289 (1-1003)

QY 220 TTCTTCGCGAGGACCTTCCCTCCAGGCAAGCGCGCGAGTTCCCGAGGAGCAG 279  
DB 1 PhePheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerLysGln 20  
QY 280 AACCGCGCCAAACAGCCACAGCGCGAGCTGCGAGTGCGCGCGACAAACCCCGCAGC 339  
DB 21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValGlnGlyAspAsnProArgSer 40  
QY 340 GAGCGCGCGCGAGCGCGAGGCGACCTTGAACTTCCCGAGATCACCTGTGCGAGCGC 399  
DB 41 GluAlaGlyValGluArgGlnGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 60  
QY 400 CCCCTGTGTAGCATCAAGGTGGCGCGCAGATCAAGGAGGCGCTGTGACACCGCGGCC 459  
DB 61 ProLeuValSerIleLysValGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80  
QY 460 GACGACACCGTGTGAGGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGGC 519  
DB 81 AspAspThrValLeuGluGluIleAlaLeuProGlyArgTrpLysProLysMetIleGly 100  
QY 520 GGCATCGCGCGCTTCATCAAGGTGGCGCGCAGTACGACGAGATCCTGTGATCGAGATCTGCGGC 579  
DB 101 GlyIleGlyGlyPheIleLysValArgGlnIleThrAspGlnIleLeuIleGluIleCysGly 120  
QY 580 AAGAAGCCCATCGGCGCGTGTGATCGGCGCCACCGCGTGAACATCATCGCGCGCAAC 639  
DB 121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsp 140  
QY 640 ATGTGTACCCAGCTGGCGTGCACCTGAATTCCTCCATCGCCCATCGAGACCGTGCCTCC 699  
DB 141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160  
QY 700 GTGAAGTGAAGCCCGCGATGAGCGGCCCAAGGTGAAGCGATGGCGCTGACCGAGGAG 759  
DB 161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180  
QY 760 AAGATCAAGCCCTGACCGCCATCTCGAGGAGTGGAGGAGGCGGCAAGTCAACCAAG 819  
DB 181 LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys 200  
QY 820 ATCGGCGCGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAAGAGGAGCAGCAC 879  
DB 201 IleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysArgLysAspSerThr 220  
QY 880 AAGTGGCGCAAGCTGTGGACTTCCGCGAGCTGAACAGCGCACCCAGGACTTCCTGGGAG 939  
DB 221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240  
QY 940 GTGAGCTGGGCATCCCGCCACCCCGCGCTGAAGAAGAAGAGACGCGTACCGTCTG 999  
DB 241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu 260  
QY 1000 GAGTGGCGGCGCTACTTTCAGGTGCTCCCTGACGAGGACTTCGCGCAAGTACACCGCC 1059  
DB 261 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheGlyLysTyrThrAla 280  
QY 1060 TTCACCATCCCGAGCATCAACAGAGACCCCGCGCATCCGCTACCAAGTACAACGTGCTG 1119  
DB 281 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 300  
QY 1120 CCCAGGCGTGAAGGCGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCCTCGAG 1179  
DB 301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320  
QY 1180 CCTTTCGCGCGCGCAACCGGAGATGTGATCTACCGAG-----GCCCGCTGTACGTG 1233  
DB 321 ProPheArgAlaArgAsnProLysIleValIleTyrGlnTyrMetAspLeuTyrVal 340

QY 1234 GGCAGCGACCTGGAGATCGGCGCAGCAGCGCGCCCAAGATCGAGGAGCTCGGCAAGCACCTG 1293  
DB 341 GlySerAspLeuGluIleGlyHisArgAlaLysIleGluGluLeuArgAlaHisLeu 360  
QY 1294 CTGCGCTGGGCTTCCACACCCCGCACAAAGACACAGAAAGAGAGCCCTTCTCTGTGG 1353  
DB 361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeuTrp 380  
QY 1354 ATGGGCTACGAGTGCACCCCGCAAGTGGACCTGCGAGCCCATCGAGCTGCCGAGAG 1413  
DB 381 MetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys 400  
QY 1414 GAGAGCTGGGCGCGTGAAGACATCCAGAGCTGTGGGCAAGCTGAACTGGCGCAGCCAG 1473  
DB 401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420  
QY 1474 ATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGCCCTG 1533  
DB 421 IleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
QY 1534 ACCGACATCTGTCCTTGCAGGAGGCGCGAGCTGTGAGCTGGCGGAGAACCCGCGAGATC 1593  
DB 441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIle 460  
QY 1594 CTGCGCGAGCCCGTGCAGCGCGTGTACTACACCCCGCAGGAGCTGGTGGCGCGAGATC 1653  
DB 461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480  
QY 1654 CAGAAGCAGGCGCCAGCAGTGGAGCTTACCAGATCTACAGAGGAGCCCTTCAAGAACCTG 1713  
DB 481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500  
QY 1714 AAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGACGCTGAAGCAGCTGACC 1773  
DB 501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520  
QY 1774 GAGCGCGTGAAGAAGTCCGATCGGAGAGCATCGTGATCTGGGCGCAAGACCCCAAGTTC 1833  
DB 521 GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp\*\*\*LysThrProLysPhe 540  
QY 1834 CGCTGCGCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGCGCAC 1893  
DB 541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAlaThr 560  
QY 1894 TGATATCCCGAGTGGGAGTTCGTGAACACACCCCGCTGTGTGAAGCTGTGTACAGCTG 1953  
DB 561 TrpIleProAspTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeu 580  
QY 1954 GAGAGAGCCCATCATCGCGCGGAGACCTTCTACGTGACGCGCGCGCCCAACCGCGAG 2013  
DB 581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600  
QY 2014 ACCAAGATCGGCGCGCGCTACGTACCGACCGCGCGCGGCGGAGAGATCGTGAGCTG 2073  
DB 601 ThrLysValGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeu 620  
QY 2074 ACCGAGACCAACCAAGAACAGACCGAGCTGCAGGCGCATCCAGCTGGCGCTCGAGGACAGC 2133  
DB 621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 640  
QY 2134 GGCAGCGAGTGAACATCTGTGACCGCAGCAGCGCATCGCCCTGGCGCATCATCCAGGCCAG 2193  
DB 641 GlyThrGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGln 660  
QY 2194 CCGCACAAGAGCGAGCGAGCTGTGAACACAGATCATCGAGCAGCTGATCAAGAAGGAG 2253  
DB 661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGlnGlnLeuIleAsnLysGlu 680  
QY 2254 AAGTGTACCTGAGCTGGTGGTGGCGCCCAAGGGCATCGCGCGCAACAGCAGCATCGAC 2313  
DB 681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysAsnGluGlnValAsp 700  
QY 2314 AAGCTGTGAGCAAGGCGCATCCCGCAAGGTGCTGTCTCTGCGACGCGCATCGAT 2364





Db 421 IletyAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
 Qy 1534 ACCGACATCGTGGCCCTGACCGAGAGAGCCGAGCTGGAGCTGGCGGAGAACCGGAGATC 1593  
 Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuLeuAlaGluAsnLysGluIle 460  
 Qy 1594 CTGGCGGAGCCGTCGACCGGCTGACTACACACCCAGCAGGACCTGGTGGCGGAGATC 1653  
 Db 461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480  
 Qy 1654 CAGAAGCAGGCGCCACGACCTGACCTACACAGATCTACACAGGACCTTCAAGAACCTG 1713  
 Db 481 GlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeu 500  
 Qy 1714 AAGACCGGCAAGTACGCCAAGATCGCACCGCCACACACGACGCTGAAGCAGCTGACC 1773  
 Db 501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520  
 Qy 1774 GAGCCCTGCAGAGATCGCCATCGAGAGCATCTGATCTGGGCAAGACCCCAAGTTC 1833  
 Db 521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTyrGlyLysIleProLysPhe 540  
 Qy 1834 CGCTGCGCCATCCAGAAGGACCTGGGAGACCTGGTGACCGACTACTGGCAGCGCCACC 1893  
 Db 541 ArgLeuProLysGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAlaThr 560  
 Qy 1894 TGGATCCCGAGTGGGAGTTCTGTAACACACCCCGCCCTGGTGAACTGTGTACAGCTG 1953  
 Db 561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTyrTyrGlnLeu 580  
 Qy 1954 GAGAAGGAGCCATCATCGGCGCCGAGACCTTCTACGTGAGCGCGCGCCCAACCGGAG 2013  
 Db 581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600  
 Qy 2014 ACCAAGATCGCAGGCGCGCTAGTGCACGACCGCGCGCGCAGAGATCGTGAGCTG 2073  
 Db 601 ThrLysLeuLysLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620  
 Qy 2074 ACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGC 2133  
 Db 621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640  
 Qy 2134 GGCAGCAGGTGAACATCGTGACCGACGACGAGTACGCCCTGGGCATCATCCAGGCCAG 2193  
 Db 641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660  
 Qy 2194 CCGACAAAGCAGCAGAGCTGTGTAACACGATCATCGACGAGCTGATCAAGAAGAG 2253  
 Db 661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGlu 680  
 Qy 2254 AAGGTGTACCTGAGCTGGGTGCCCGCCCAAGGGCATCGCGGCAACGAGCAGATCGAC 2313  
 Db 681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnValAsp 700  
 Qy 2314 AAGCTGTGAGCAAGGCGATCCCGAAGGTGTGTCTCTGACGCGCATCGAT 2364  
 Db 701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 6  
 AAB69287  
 ID AAB69287 standard; protein; 1005 AA.  
 XX  
 AC AAB69287;  
 AC  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 DE HIV-1 non-subtype B clone 962W751-3 pol protein.  
 XX  
 KW HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 XX  
 OS Human immunodeficiency virus 1.

XX WO200026416-A1.  
 PN 11-MAY-2000.  
 PD 25-OCT-1999; 99WO-US024837.  
 PP 02-NOV-1998; 98US-00184418.  
 PR (UABR-) UAB RES FOUND.  
 PA Hahn BH, Shaw GM, Gao F;  
 PI WPI; 2000-365651/31.  
 DR Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 XX type 1 useful for detecting and treating AIDS comprises a specific  
 XX nucleotide sequence.  
 PS Claim 41; Fig 15; 131pp; English.  
 CC The present invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1005 AA;  
 Alignment Scores:  
 Pred. No.: 5e-200 Length: 1005  
 Score: 3605.00 Matches: 681  
 Percent Similarity: 96.68% Conservative: 18  
 Best Local Similarity: 94.19% Mismatches: 16  
 Query Match: 78.85% Indels: 8  
 DB: 3 Gaps: 2  
 US-09-610-313B-31 (1-2463) x AAB69287 (1-1005)

Qy 220 TTCTTCGCGAGACCTGGCGCTTCCCGCAGGCAAGGCCCGCGAGTTCCTCCGAGGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20  
 Qy 280 AACCGCGCC-----AACAGCCCGCCAGCCCGCGAGCTCGAGTGCAGGTGCGC 321  
 Db 21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40  
 Qy 322 GCGACAAACCCCGCAGCG 381  
 Db 41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60  
 Qy 382 ATCACTCTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGCGCGCGCGCGCGCGCGCGCG 441  
 Db 61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIleLysGluAla 80  
 Qy 442 CTGCTGGACACCG 501  
 Db 81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrp 100  
 Qy 502 AAGCCCAAGATGATCG 561  
 Db 101 LysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIle 120  
 Qy 562 CTGATCGAGATCTCGCGCAAGAAGCGCATCGGACCGCTGTGTGTGTGTGTGTGTGTGT 621  
 Db 121 LeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVal 140  
 Qy 622 AACATCATCG 681  
 Db 141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160

```
QY 682 CCCATCGAGACCGTGCCTGAAGCTGAAGCCCGCATGACCGCCCAAGGTGAAGCAG 741
DB |||||
DB 161 ProIleGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln 180
QY 742 TGGCCCTGTACCGAGAGAAAGATCAAGGCCCTGACCGCCATCTCGGAGGAGATGGAAG 801
DB |||||
DB 181 TrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys 200
QY 802 GAGGGCAAGATCACCAAGATCGCGCCCGGAGAACCCCTACCAACCCCGTGTGCGCATC 861
DB |||||
DB 201 GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle 220
QY 862 AAGAAGAGACAGACCAAGTGGCGCAAGCTGGTGACTTCGCGCAGCTGGAACAAGCGC 921
DB |||||
DB 221 LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArg 240
QY 922 ACCGAGGACTTCTGGGAGTGTGAGCTGGGCATCCCCACCCCGCCGCTGGAAGAAG 981
DB |||||
DB 241 ThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLys 260
QY 982 AAGAGGTGACCGTGTGAGCTGGGCGGAGCCCTACTTACGCTGCGCCCTGGACGAGGAC 1041
DB |||||
DB 261 LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly 280
QY 1042 TTCGCGAAGTACACCGCTTCACCATCCCGCAGCATCAACAACGAGACCCCGCCGATCCGC 1101
DB |||||
DB 281 PheArgLysIleThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyLysArg 300
QY 1102 TACCAGTACAACTGTGTCGCCCGAGGCTGGAAGGGCAGCCCGCAGCATCTTCCAGAGCAGC 1161
DB |||||
DB 301 TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer 320
QY 1162 ATGACCAAGATCTGTGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGACTACCGAG --- 1218
DB |||||
DB 321 MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr 340
QY 1219 ---GCCCCCTGTACGTGGGCGAGCGACTCGAGATCGGCAGCACCGCGCAAGATCGAG 1275
DB |||||
DB 341 MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu 360
QY 1276 GAGCTGCGCAAGCACTGTGCTGCGCTGGGGTTCCACCAACCCCGCAAGAGCAAGCAAG 1335
DB |||||
DB 361 GluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLys 380
QY 1336 GAGCCCCCTCTCTGTGGATGGGCTAGAGCTGCACCCCGCAGTGGAGCGTGCAGCCC 1395
DB |||||
DB 381 GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnPro 400
QY 1396 ATCGAGTCCCGGAGAGAGAGCTGGACCGTGAACGACATCCAGAACTGGTGGGCAAG 1455
DB |||||
DB 401 IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys 420
QY 1456 CTGAACCTGGCGCAGCCAGATCTACCCCGGCATCAAGGTGCGCAGCTGTGCAAGCTGCTG 1515
DB |||||
DB 421 LeuAsnTrp**SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu 440
QY 1516 CGCGCGCCCAAGGCCCTGACCGCATCTGCGCCCTGACCGAGAGCGCGCTGGAGCTG 1575
DB |||||
DB 441 ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeu 460
QY 1576 GCCGAGAACCGCAGATCTCTGGCGAGCCCGTGCACCGCGTGTACTACGACCCCGCAGCAAG 1635
DB |||||
DB 461 AlaGluSerArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLys 480
QY 1636 GACCTGTGTGCGCAGATCTCCAGAACGAGGGCCACGACAGTGGACCTTACAGATCTACCA 1695
DB |||||
DB 481 AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln 500
QY 1696 GAGCCCTTCAAGACCTGAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACCAAC 1755
DB |||||
DB 501 GluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsn 520
```

```
QY 1756 GACGTCAAGCAGCTGATCCGAGGCCGTGTCAGAGATCGCCATGAGAGCATCTGTGATCTGG 1815
DB |||||
DB 521 AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrp 540
QY 1816 GGCAGACACCCCAAGTTCGCTGCGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACC 1875
DB |||||
DB 541 GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr 560
QY 1876 GACTACTGGCAGGCCACCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTG 1935
DB |||||
DB 561 AspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal 580
QY 1936 AAGCTGTGTACAGCTGGAGAGAGCCCATCATCGGCCCGCAGACCTTCTACGTGGAC 1995
DB |||||
DB 581 LysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp 600
QY 1996 GGGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGGTACGTGACCCGACCCGGGCGCG 2055
DB |||||
DB 601 GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg 620
QY 2056 CAGAAGATCGTGAGCCTGACCGAGACCAACCAACAGAACCCGAGCTGCAGGCCATCCAG 2115
DB |||||
DB 621 GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln 640
QY 2116 CTGGCCCTCGCAGACAGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTG 2175
DB |||||
DB 641 LeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeu 660
QY 2176 GGCAATCATCCAGGCCCGCCGACAGAGCGGAGAGCTGGTGAACAGATCATTCGAG 2235
DB |||||
DB 661 GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGlu 680
QY 2236 CAGCTCATCAAGAAGAGAGAGGTGTACTCGAGCTGGGTGCCCGCCACAGGGCATCGGC 2295
DB |||||
DB 681 GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly 700
QY 2296 GGCACAGCAGCAGATCCACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCTGGAC 2355
DB |||||
DB 701 GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp 720
QY 2356 GGCATCGAT 2364
DB |||||
DB 721 GlyIleAsp 723
RESULT 7
AAE37601
ID AAE37601 standard; protein; 998 AA.
XX
AC AAE37601;
XX
XX 23-OCT-2003 (revised)
DT 27-AUG-2003 (first entry)
XX
DE HTV-1 subtype C isolate Du151 reverse transcriptase (RT) protein.
XX
XX Regulatory gene; accessory gene; HIV; human immunodeficiency virus;
XX vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.
XX
XX Human immunodeficiency virus 1.
OS
XX WO2003037919-A2.
PN
XX
XX 08-MAY-2003.
PD
XX
XX 31-OCT-2002; 2002WO-1B004550.
PF
XX
XX 31-OCT-2001; 2001ZA-00008978.
PR
XX
XX (SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.
PA (UYCA-) UNIV CAPE TOWN.
XX
XX Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;
PI
XX
```

DR WPI: 2003-430497/40.  
 XX N-PSDB; AAD29258.  
 PT New molecules comprising HIV-1 subtype isolate regulatory/accessory  
 PT genes, useful for manufacturing a vaccine for treating or preventing HIV  
 PT infection.  
 XX Disclosure; Page 94-97; 97pp; English.

CC The invention relates to molecules comprising HIV-1 subtype isolate  
 CC regulatory/accessory genes (tat, nef and rev genes) and modifications and  
 CC derivatives thereof. The invention also provides proteins encoded by such  
 CC genes. Sequences of the invention are useful for manufacturing vaccines  
 CC for treating or preventing human immunodeficiency virus (HIV) infections.  
 CC They are also useful in gene therapy. The present sequence is HIV-1  
 CC subtype C isolate Du151 reverse transcriptase (RT) protein. (Updated on  
 CC 23-OCT-2003 to standardise OS field)

XX Sequence 998 AA;  
 SQ

Alignment Scores:  
 Pred. No.: 6.01e-198 Length: 998  
 Score: 3570.00 Matches: 666  
 Percent Similarity: 97.07% Conservative: 29  
 Best Local Similarity: 93.02% Mismatches: 19  
 Query Match: 78.07% Indels: 2  
 DB: 6 Gaps: 1

US-09-610-313B-31 (1-2463) x AAE37601 (1-998)

QY 223 TTCCGCGAGGACCTGGGCTTCCCGCAGGCGGAGCCCGCGAGTTCCCGAGGACGAGAC 282  
 DB 1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20  
 QY 283 CGGCGCAACAGCCACAGCGGAGCTGCAGGTGGCGCGGACAAACCCCGCAGCGAG 342  
 DB 21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgAsnProArgSerGlu 40  
 QY 343 GCCGCGCGAGCGCCAGGCGACCTGAACCTCCCGCAGATCACCTGTGGCAGCGCCC 402  
 DB 41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnLeuThrLeuTrpGlnArgPro 60  
 QY 403 CTGGTGAAGTCAAGTGGGCGGCACATCAGAGAGCCCTGTGGACACCGGCGCGAC 462  
 DB 61 LeuValSerIleLysIleGlyGlyThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80  
 QY 463 GACACCGTGTGGAGGAGATGAGCTCCCGGCAAGTGAAGCCCAAGATGATCGGCGGC 522  
 DB 81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100  
 QY 523 ATCGCGCGCTTCATCAAGGTGCGCCAGTACGACACAGATCCTGTATCGAGATCTGCGGCAAG 582  
 DB 101 IleGlyGlyPheIleLysValArgGlnThrAspGlnLeuLeuIleGluIleCysGlyLys 120  
 QY 583 AAGGCCATCGGCAAGCTGTGATCGGCGCCACCCCGCTGAACATCATCGGCGCGCAACATG 642  
 DB 121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140  
 QY 643 CTGACCCAGCTGGGCTGCACCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGGT 702  
 DB 141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160  
 QY 703 AAGCTGAAGCCCGCATGAGCGGCCCAAGGTGAAGCAGTGGCCCTTACCGAGGAGAAG 762  
 DB 161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluLys 180  
 QY 763 ATCAAGCCCTGACCGCATCTGCGAGGAGATGAGAGGAGGCGCAAGATCACCAAGATC 822  
 DB 181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200  
 QY 823 GGGCCCGGAGACCCCTACACACCCCGCTGTCGCATCAAGAGAGAGGACAGCACCAAG 882  
 DB 201 GlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

QY 883 TGGCGCAAGCTGTGGAGCTTCCCGAGAGTGAAAGCGCACCCAGGACTTCTTGGAGGTG 942  
 DB 221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240  
 QY 943 CAGCTGGGCATCCCGCACCCCGCGCTGAAGAAGAAGAGAGCGTGACCGTGTCTGGAC 1002  
 DB 241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAsp 260  
 QY 1003 GTGGCGCAGCCTACTTTCAGCGTCCCGCTGACGAGGACTTCCGCAAGTACACGCGCTTC 1062  
 DB 261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280  
 QY 1063 ACCATCCCGAGCATCAACAGAGAGACCCCGCGCATCCGCTACCAAGTACAACGTCGTGCC 1122  
 DB 281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300  
 QY 1123 CAGGCGTGAAGGCGCGCCAGCATCTTCAGAGCAGCAGTACGACAGATCCTCGAGCGCC 1182  
 DB 301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320  
 QY 1183 TTCCGCGCGCGCAACCCCGAGATCGTGATCTTACCAG-----GCCCGCTGTAGCTGGGC 1236  
 DB 321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340  
 QY 1237 AGCGACCTGAGATCGGCGCAGCACCGCGCCCAAGATCAGAGAGCTGCGCAAGCAGCTGCTG 1296  
 DB 341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360  
 QY 1297 CGCTGGGCGTTCACCCCGCAACAGCAGCAGAGGAGCGCCCTTCTCTGTGGATG 1356  
 DB 361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380  
 QY 1357 GGCTTACGAGCTGCACCCCGCAAGTGCAGCGCTGCAGCCCATCGAGCTGCCCGAAGAGAG 1416  
 DB 381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400  
 QY 1417 AGCTGAGCCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAACTGGGCGCAGCAGATC 1476  
 DB 401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420  
 QY 1477 TACCCCGGCACTAAGTGGCGCGAGCTGTGCAAGCTGTGCGCGCGCGCCAGGCGCTGACC 1536  
 DB 421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440  
 QY 1537 GACATCTGCTGCGCTGACCGAGGAGCGAGCTGAGCTGGCGGAGAACCCGAGACTCTG 1596  
 DB 441 AspIleValProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeu 460  
 QY 1597 CGCGAGCCCTGCAAGCGCTGTACTACGACCCCAAGCAAGACCTGTGTGGCGGAGATCCAG 1656  
 DB 461 LysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGln 480  
 QY 1657 AAGCAGGCGCACCGACCTGAGCTTACAGATCTTACAGGAGCGCTTCAAGACCTGAG 1716  
 DB 481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500  
 QY 1717 ACCGCGCAAGTACGCGCAAGTGCAGCGCCCGCACACCAACACGACGTAAGAGCAGCTCACCGAG 1776  
 DB 501 ThrGlyLysTyrAlaLysArgArgThrHisThrAsnAspValLysGlnLeuThrGlu 520  
 QY 1777 GCCGTGCAAGATCGCCATGGAGAGCATCGTGATCTGGGGCAAGACCCCAAGTTCGCG 1836  
 DB 521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysTrpProLysPheArg 540  
 QY 1837 CTGCCCATCCAGAGGAGACCTGGGAGACCTGTGGAGCGGACTACTGGCGGCGCACCTGG 1896  
 DB 541 LeuProIleGlnLysGluThrTrpGluIleTrpThrAspTyrTrpGlnAlaThrTrp 560  
 QY 1897 ATCCCGGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGTGTGCTACCACTGGAG 1956  
 DB 561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580



QY 723 CGGCCCCAAGGTGAAGCAAGTGGCCCTCTGACCCGAGGAGAGATCAAGGCCCTGACCGCCAT 782  
 Db 654 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIle 674  
 QY 783 CTGGGAGAGATGAGAGAGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTACAA 842  
 Db 674 eCyAsnGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAs 694  
 QY 843 CACCCCGTGTTCGCCATCAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902  
 Db 694 nThrProIlePheAlaIleLysLysAspSerThrLysIleAspValLeuValAspPh 714  
 QY 903 CGGAGAGTGAACAGCCGACCCAGGACTTCTGGAGGTGAGTGGGATCCGCCACCC 962  
 Db 714 eArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734  
 QY 963 CGCCGGCTGAAG 1022  
 Db 734 oAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754  
 QY 1023 CGTGGCCCTGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCCGCAGCATCAACA 1082  
 Db 754 rIleProLeuThrGluAspPheArgLysThrAlaPheThrIleProSerArgAsnAs 774  
 QY 1083 CGAGACCCCGGATCCGCTACAGTACAGTGTGCTGCTGCCCGGAGGAGGAGGAGGAG 1142  
 Db 774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGlySerLe 794  
 QY 1143 CAGCATCTTCAGAGCAGCATGACCAAGATCTCGAGGCCCTTCGCGGCCCGCAGCCCGA 1202  
 Db 794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814  
 QY 1203 GATCGTGTATACAG-----GCCCGCTGTAGCTGGGAGGAGGAGGAGGAGGAGGAG 1256  
 Db 814 yIleValIleThrGlnTyrMetAspLeuThrValGlySerAspLeuGluIleGlyGl 834  
 QY 1257 GCACCGGCGGAGATGAGAGAGTGGCAAGCAGTCTGCTGGCTGGGCTTCCACACCCC 1316  
 Db 834 nHisArgThrLysIleGluLeuArgLysGlnHisLeuLeuArgTrpGlyPheThrThrPr 854  
 QY 1317 CGACAAGAGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1376  
 Db 854 oAspLysLysHis---LysGluProProPheLeuThrMetGlyTyrGluLeuHisProAs 873  
 QY 1377 CAAGTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1436  
 Db 873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerThrValAsnAspIl 893  
 QY 1437 CCAGAAGCTGTGGCAAGTGAAGTGGCCAGGAGTACCCCGGAGTCAAGGTGCG 1496  
 Db 893 eGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValAr 913  
 QY 1497 CCAGCTGCAAGCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1556  
 Db 913 gGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933  
 QY 1557 GGAGCGGAGCTGAGCTGGCGGAGAACCGGAGATCTGCGGAGGCGGCGGCGGCGGCGG 1616  
 Db 933 uGluAlaGluLeuLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVa 953  
 QY 1617 GTACTAGACCCGAGCAGGAGTGTGGTGGCGAGATCCAGAAGCAGGAGGAGGAGGAGG 1676  
 Db 953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973  
 QY 1677 GACCTACAGATCTACAGAGAGGCGCTTCAAGACCTGAAGACCGGCGAGTACGCGCAAG 1736  
 Db 973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993  
 QY 1737 GCGCAGCGCCACACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1796  
 Db 993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

QY 1797 GGAGAGCATCTGATCTGGGCAAGACCCCAAGTTCCGCTGCGCTCCATCCAGAGAGAC 1856  
 Db 1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033  
 QY 1857 CTGGGAGACCTGGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTCGT 1916  
 Db 1033 rTrpGluThrTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053  
 QY 1917 GAACACCCCGCTGGTGAAGCTGTGTATCCAGCTGGAGAGAGGAGGAGGAGGAGGAG 1976  
 Db 1053 lAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysAspProIleValGlyVa 1073  
 QY 1977 CGAGACCTTCTACGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTA 2036  
 Db 1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTy 1093  
 QY 2037 CGTGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2096  
 Db 1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113  
 QY 2097 CGAGCTGCGAGGCCATCCAGCTGGCCCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 2156  
 Db 1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133  
 QY 2157 CGAGACGAGTACGCTGGGATCATCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2216  
 Db 1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153  
 QY 2217 GGTGAACCGAGTATCCGAGCAGCTGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 2276  
 Db 1153 uValAsnGlnIleIleGluGlnLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173  
 QY 2277 CGCCCAAGGCGATCGGCGGCAACGAGCAGATCGACAAGCTGGTGGTGAAGAGGCGATCCG 2336  
 Db 1173 oAlaHisLysGlyIleGlyLysAsnGluGlnValAspLysLeuValSerSerGlyIleAr 1193  
 QY 2337 CAAGGTGCTGCTTCGAGCGGATCGAT 2364  
 Db 1193 gLysValLeuPheLeuAspGlyIleAsp 1202

RESULT 9  
 AAM48949  
 ID AAM48949 standard; protein; 854 AA.  
 XX  
 AC AAM48949;  
 XX  
 DT 29-AUG-2003 (revised)  
 DT 19-APR-2002 (first entry)  
 XX  
 DE HIV-1 subtype C isolate Dul51 pol protein.  
 XX  
 KW HIV-1 subtype C; vaccine; HIV infection; AIDS; pol; antiviral.  
 XX  
 OS Human immunodeficiency virus; type I.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 58 /note= "encoded by TGA"  
 FT Misc-difference 72 /note= "encoded by TAA"  
 FT Misc-difference 114 /note= "encoded by TAA"  
 FT Misc-difference 117 /note= "encoded by TAA"  
 XX  
 FN WO200204494-A2.  
 XX  
 PD 17-JAN-2002.  
 XX  
 PF 09-JUL-2001; 2001WO-IB001208.  
 XX  
 PR 07-JUL-2000; 2000US-0216995P.  
 PR 10-JUL-2000; 2000ZA-00003437.

```
PR 15-SEP-2000; 2000ZA-00004924.
XX (MEDI-) MEDICAL RES COUNCIL.
PA (UYCA-) UNIV CAPE TOWN.
PA (UYNC-) UNIV NORTH CAROLINA.
XX
PI Williamson C, Swannstrom RI, Morris L, Karim SA, Johnston RE;
XX WPI; 2002-171700/22.
DR N-PSDB; AAL41592.
XX
XX Selecting HIV-1 subtype C isolates, which are useful in developing
PT vaccines against HIV infection, comprises isolating viruses with high
PT sequence identity to a consensus sequence whose phenotype is associated
PT with the HIV subtype.
XX
XX Claim 19; Page 67; 69pp; English.
XX
XX The present invention relates to a process for the selection of human
CC immunodeficiency virus (HIV) subtype isolates for use in the development
CC of a prophylactic and/or therapeutic pharmaceutical composition. The
CC process involves selecting isolated virus or viruses with a high sequence
CC identity to a consensus sequence and a phenotype which is associated with
CC transmission for the particular HIV subtype. The composition can be used
CC in the production of vaccines against HIV. The present sequence is the
CC HIV-1 subtype C isolate Du151 pol protein. (Updated on 29-AUG-2003 to
CC standardise OS field)
XX
XX SQ Sequence 854 AA;

Alignment Scores:
Pred. No.: 3 94e-195 Length: 854
Score: 3521.00 Matches: 666
Percent Similarity: 94.31% Conservatives: 30
Best Local Similarity: 90.24% Mismatches: 29
Query Match: 77.00% Indels: 13
DB: 5 Gaps: 3

US-09-610-313B-31 (1-2463) x AAM48949 (1-854)
QY 171 CGCAAGGAGGCGCCACGATGAAGACTGCACCGAGCGCCAGGCCAACTCTT----- 224
Db 118 ArgGlnGlyPheProSerHisAppValValLysArgArgProValProSerLeuHisAla 137
QY 225 -----CGCGAGGACCTGGCC--TTCCCGCGGCGAAGCGCCG 260
Db 138 CysArgSerThrLeuGluAppProArgValPro-SerSerPheProGlnGlyProAla 157
QY 261 CGAGTTCCTCCAGGAGCAGAACCGCGCAACAGCCCGCCAGCGCGCGCGAGCTGCGGCG 320
Db 157 GlnPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnVal 177
QY 321 CGCGCGAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
Db 177 GArgAspAsnProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAsnPhePro 197
QY 381 GATCACCTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
Db 197 nllleThrLeuTrpGlnArgProLeuValSerlleLyslleGlyGlyGlnThrArgGlu 217
QY 441 CTGCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
Db 217 aLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspilleAsnLeuProGlyLys 237
QY 501 GAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
Db 237 pLysProLysMetileGlyGlyGlyGlyGlyPheLysValArgGlnThrArgGln 257
QY 561 CTGATCGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620
Db 257 eLeuileGluileCysGlyLysLysAlaileGlyThrValLeuValGlyProthPro 277
QY 621 GAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTCGAATCTCCCATCAG 680

277 lAsnilleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSe 297
681 CCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGGAGCGGCCCAAGGTGAAGCA 740
297 rProfileGluThrValProValLysLeuLysProGlyMetAspGlyProLysValysG 317
741 GTGGCCCTGACCGAGGAGAGATCAAGGCCCTTACCGCCATCTGCGAGAGAGATGAGAA 800
317 nTrpProLeuThrGluValLyslleLysAlaLeuThrAlaileCysGluGluMetGlu 337
801 GAGGCGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTTCAACACCCCGTGTTCGCCAT 860
337 sGluGlyLyslleThrLysileGlyProGluAsnProTyrAsnThrProIlePheAla 357
861 CAAGAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCCGGAGGTGAACAGCG 920
357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 377
921 CACCCAGGACTTCTGGGAGGTGCAGCTGGGCGATCCCGCCACCCCGCGCGCTGAAGAA 980
377 gThrGlnAspPheTrpGluValGlnLeuGlylleProHisProAlaGlyLeuLysLys 397
981 GAAGACCGTGAACCGTCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040
397 sLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGlu 417
1041 CTTCCGCAAGTACACCGCTTACCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1100
417 yPheArgLysTyrThrAlaPheThrileProSerlleAsnAsnGluThrProGlylle 437
1101 CTACCAAGTACAACTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1160
437 gTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlailePheGlnAla 457
1161 CATGACCAAGATCCTCGAGCCCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1218
457 rMetThrLyslleLeuGluProPheArgAlaLysAsnProGluileValleLysrGln 477
1219 -----GCCCGCTGTAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1274
477 rMetAlaAlaLeuTyrValGlySerAspLeuGluileGlyGlnHisArgAlaLyslle 497
1275 GAGCTGCGCGCAAGCACCTGCTGCGCTGGGGCTTTCACACACCCCGCGCAAGAACCA 1334
497 uGluLeuArgGluHisLeuLysTrpGlyPheThrThrProAspLysLysHisGln 517
1335 GAGCGCGCGCTTCTCTGTGATGGGCTACGAGCTGCGCGCGCGCGCGCGCGCGCGCG 1394
517 sGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGln 537
1395 CATCGAGCTGCCGAGAGGAGCTGGACCGTGAACGACATCCAGAGCTGTGTGGGCAA 1454
537 oilleGlnLeuProGluLysAspSerTrpThrValAsnAspilleGlnLysLeuVal 557
1455 GCTGAATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1514
557 sLeuAsnTrpThrSerGlnlleTyrProGlylleLysValArgGlnLeuCysLysLe 577
1515 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1574
577 uArgGlyThrLysAlaLeuThrAspilleValProLeuThrGluLalaGluLeuGlu 597
1575 GCGCGAGACCGCGAGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1634
597 uAlaGluAsnArgGluileLeuLysGluProValHisGlyValTyrTrpAspProSer 617
1635 GGAACCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1694
617 sAspLeuileAlaGluileGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 637
1695 GAGCGCGCTTCAAGAACTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1754
```

Db 637 nGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysArgArgThrThrHisThrAs 657  
Qy 1755 CGAGCTGAAGCAGCTGACGCGGCGCTGCAGAGATCCGATCGAGAGCATCGTGATCTG 1814  
Db 657 nAspValLysGlnLeuThrGluAlaValGlnLysLeuSerLeuGluSerLeuValThrTr 677  
Qy 1815 GGGCAAGACCCCAAGTTCGCTGCCCTGCCATCCAGAGGAGACTGGGAGACTGGTGAC 1874  
Db 677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluIleThrTrpTh 697  
Qy 1875 CGACTACTGGCAGCCCACTGGATCCCGAGTGGAGTTCGTGAACACCCGCCCTGGT 1934  
Db 697 rAspTyrTrpGlnAlaThrTrpIleProGluThrPheValAsnThrProProLeuVa 717  
Qy 1935 GAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGGCGCCGAGACTTCTACGTGGA 1994  
Db 717 lLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPheTyrValAs 737  
Qy 1995 CGGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGCGCG 2054  
Db 737 pGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyAr 757  
Qy 2055 GCAGAAGATCTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCA 2114  
Db 757 gGlnLysIleValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGl 777  
Qy 2115 GCTGCGCCTCGAGACAGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCT 2174  
Db 777 nLeuAlaLeuGlnAspSerGluSerGluValAsnIleValThrAspSerGlnTyrAlaLe 797  
Qy 2175 GGGCATATTCAGGCCCGCCGACAGAGCGAGCGAGCGAGCTGTGTGAACCGATCATCGA 2234  
Db 797 uGlyIleIleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnIleIleGl 817  
Qy 2235 GCAGCTCATCAAGAGGAGAGGTGTACCTGAGCTGGTGCCTCCGCCACAGGCGCATCGG 2294  
Db 817 uGlnLeuLysLysGluArgAlaTyrLeuSerTrpValProAlaHisLysGlyIleGl 837  
Qy 2295 CGGCAAGCAGCAGATGACAGCTGGTGGAGCAAGCGGATCGCGAAGGTGCTG 2346  
Db 837 yGlyAspGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 854  
RESULT 10  
AAB69282  
ID AAB69282 standard; protein; 1000 AA.  
XX  
AC AAB69282;  
XX  
XX  
DT 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX  
XX HIV-1 non-subtype B clone 92RW009-6 pol protein.  
XX  
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
XX vif; vpr; tat; rev; nef; vaccine.  
XX  
XX Human immunodeficiency virus 1.  
XX  
XX WO20026416-A1.  
XX  
XX 11-MAY-2000.  
XX  
XX 25-OCT-1999; 99WO-US024837.  
XX  
XX 02-NOV-1998; 98US-00184418.  
XX  
XX (UABR-) UAB RES FOUND.  
XX  
XX Hahn BH, Shaw GM, Gao F;  
XX  
XX WFI; 2000-365651/31.  
XX  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX  
PS Claim 41; Fig 15; 131pp; English.  
XX  
CC The present in invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX

Sequence 1000 AA;

Alignment Scores:  
Pred. No.: 6,5e-195 Length: 1000  
Score: 3517.50 Matches: 660  
Percent Similarity: 95.96% Conservative: 29  
Best Local Similarity: 91.92% Mismatches: 26  
Query Match: 76.92% Indels: 3  
DB: 3 Gaps: 2

US-09-610-313B-31 (1-2463) x AAB69282 (1-1000)

Qy 220 TTCTTCGCGAGGACCTGGCTTCCCCAGGCAAGCGCCGCGAGTTCCCCAGCGAGCAG 279  
Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20  
Qy 280 AACCGCCGCAACACCCGCCAGCCGCGAGCTGCAGGTGCGCGCGC---GACACCCCGCGC 336  
Db 21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40  
Qy 337 AGCGAGCGCGCGCGAGCGCGAGCGCACCTGAATTTCCCCAGATCACCCTGTGGCAG 396  
Db 41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnIleThrLeuTrpGln 60  
Qy 397 CGCCCCCTGTGTGATCAAGGTGGCGCGCGCAGATCAAGGAGGCGCCTGCTGGACACCGCGC 456  
Db 61 ArgProLeuValThrValLysIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80  
Qy 457 GCGACGACACCGCTGTGGAGGAGATGAGCTGCCCGCGCAAGTGAGAGCCCAAGATGATC 516  
Db 81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMethIle 100  
Qy 517 GCGGCGATCGCGCGCTTCATCAGGTGGCGCGCAGTACGACGAGATCCTGTGATCGATCTGC 576  
Db 101 GlyGlyIleGlyGlyPheIleLysValLysGlnTyrAspGlnIleLeuIleGluIleCys 120  
Qy 577 GCGAAGAGCGCCATCGCGCACCGCTGTGATCGGCGCGCACCCCGCTGAACATCATCGCGCGC 636  
Db 121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleGlyArg 140  
Qy 637 AACATGTGACCGAGCTGGGCTGCACCTGAACTTCCCATCAGCCCATCGACACCGGTG 696  
Db 141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160  
Qy 697 CCGTGAAGCTGAAGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 756  
Db 161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTyrProLeuThrGlu 180  
Qy 757 GAGAAGATCAAGCGCGCTGACCGCCATCTGCGAGGAGATGGAAGAGGAGGGCAAGATCAC 816  
Db 181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluLysLysIleSer 200  
Qy 817 AAGATCGCGCGCGAGAACCCCTACAACACCCCGCTGTTCGCCATCAGAGAGAGCAGCAGC 876  
Db 201 LysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSer 220  
Qy 877 ACCAAGTGGCGAGCTGTGGACTTCGCGAGCTGAACAGCGCGCGCGCGCGCGCGCGCGCG 936  
Db 221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240



QY 937 GAGTGCAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAAGACGCGTGACCGTG 996  
Db 241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVal 260  
QY 997 CTGGACCTGGCGCAGCGCTACTTTCAGCGTCCCTCGACGAGGACTTCGCAAGTACACC 1056  
Db 261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280  
QY 1057 GCCTTCAATCCCGACATCAACACGAGACACCCCGGCATCCGCTACGATCAACAGTG 1116  
Db 281 AlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVal 300  
QY 1117 CTGCCCCAGCGTGGAGGCGACCCCGACGATCTTCCAGAGCAGCATGACCAAGATCCTG 1176  
Db 301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320  
QY 1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCTACACAG-----GCCCCCTGTAC 1230  
Db 321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspAspLeuTyr 340  
QY 1231 GTGGGACGACCTGGAGATCGCGCAGCACCGCGCCMAAGATCGAGAGCTGCGCAAGCAC 1290  
Db 341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHis 360  
QY 1291 CTGCTGCGCTGGGCTTCCACACCCCGCACAGAGCACGAGAGGAGCCCTTCTCCTG 1350  
Db 361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeu 380  
QY 1351 TGGATGGGCTTACGAGCTGCACCCCGACCAAGTGACCGTGCAGCCCATCGAGCTGCCCGAG 1410  
Db 381 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400  
QY 1411 AAGAGAGCTGGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACCTGGCGCAGC 1470  
Db 401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420  
QY 1471 CAGATCTACCCCGCATAGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCCAGAGCC 1530  
Db 421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440  
QY 1531 CTGACCGACATCTGTCGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAG 1590  
Db 441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGlu 460  
QY 1591 ATCTGGCGCAGCCGTGCACGCGTGTACTAGACCCCGACGAGACCTGTGGCGCGAG 1650  
Db 461 IleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGlu 480  
QY 1651 ATCCAGAAGCAGGCGCCACGACGAGTGGACCTTACCAGATCTACCAGGAGCCCTTCAAGAAC 1710  
Db 481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500  
QY 1711 CTGAAGACCGGCAAGTACGCGCAAGATCGCGCACCGCCACACCAACGACGCTGAAGACGCTG 1770  
Db 501 LeuLysThrGlyLysTyrAlaLysArgArgThrAlaHisThrAsnAspValLysGlnLeu 520  
QY 1771 ACCGAGCCGTGCAAGAGATCGCCATGGAGAGCATCTGTATCTGGGGCAAGACCCCAAG 1830  
Db 521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrpGlyLysThrProLys 540  
QY 1831 TTCGCTGCTCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGGCAGGCC 1890  
Db 541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTrpTrpGlnAla 560  
QY 1891 ACCTGATCCCGAGTGGGAGTTCTGTGAACACCCCGCTGTGTGAGCTGTGTATCCAG 1950  
Db 561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580  
QY 1951 CTGAGAAGGAGCCCATCATCTGGCGCGCGAGACCTTCTACGTGAGCGCGCGCCCAACCGC 2010  
Db 581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

QY 2011 GAGACCAAGATCGCGCAAGGCCGGCTACGTACCGACCGCGCGCGCGAGAGATCGTGAGC 2070  
Db 601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620  
QY 2071 CTGACCGAGACCAACCAAGAACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGAC 2130  
Db 621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640  
QY 2131 AGCGGCGAGCGGTGAACATCGTGACCGACGACGAGCTAGCCCTGGGCATCATCCAGGCC 2190  
Db 641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 660  
QY 2191 CAGCCCGCAAGACGAGCGAGCGAGCTGTGTAACACGAGATCATCGAGCAGCTGATCAAGAAG 2250  
Db 661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleIleGlnLeuLysLys 680  
QY 2251 GAGAAAGTGTACGTAGCTGGGTGCGCGCCCAAGGGCATCGCGGCGCAACGAGCAGATC 2310  
Db 681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyAsnGluGlnVal 700  
QY 2311 GACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCGACGCGCATCGAT 2364  
Db 701 AspLysLeuValSerSerGlyIleArgValLeuPheLeuAspGlyIleAsp 718  
RESULT 11  
AAW72993  
ID AAW72993 standard; protein; 1002 AA.  
XX AC AAW72993;  
XX XX 25-MAR-2003 (revised)  
DT 15-FEB-1999 (first entry)  
XX XX HIV isolate LAV.MAL pol protein.  
DE LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;  
KW pol.  
XX Human T-lymphotropic virus.  
OS  
XX US5824482-A.  
FN 20-OCT-1998.  
XX XX 06-JUN-1995; 95US-00471474.  
XX XX 23-JUN-1986; 86FR-00040138.  
PR 13-APR-1987; 87US-00038330.  
PR 19-FEB-1991; 91US-00656797.  
PR 10-DEC-1992; 92US-00988530.  
PR 18-NOV-1993; 93US-00154397.  
XX (INSP ) INST PASTEUR.  
PA Wain-Hobson S, Sonigo P, Alizon M, Montagnier L;  
XX WPI: 1998-582548/49.  
XX N-PSDB; AAV63467.  
XX Human immunodeficiency virus isolate LAV(MAL) - and method for detecting  
PT anti-HIV antibodies.  
PS Disclosure; Fig 7B-E; 47pp; English.  
XX This is the amino acid sequence of the pol protein of lymphadenopathy  
CC associated virus LAV.MAL (CNCM 1-641), a new virus isolate from Zaïre  
CC that is responsible for diseases clinically related to AIDS. The sequence  
CC was deduced from an open reading frame (ORF) of the LAV.MAL genome (see  
CC AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of  
CC the envelope glycoprotein can be used as antigens in a claimed method for  
CC the in vitro detection of an antibody directed against LAV. The method is  
CC useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in  
CC patients, asymptomatic carriers and in blood-related products. LAV viral

CC antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct PR field.)

XX  
SQ Sequence 1002 AA;

Alignment Scores:  
Pred. No.: 7,14e-194 Length: 1002  
Score: 3499.50 Matches: 653  
Percent Similarity: 96.11% Conservatism: 39  
Best Local Similarity: 90.69% Mismatches: 23  
Query Match: 76.53% Indels: 5  
DB: 2 Gaps: 3

US-09-610-313B-31 (1-2463) x AAU72993 (1-1002)

QY 220 TTCTTCGGCAGGACCTGGCCCTTCCCGCAGGCGAAGGCCCGCGAGTTCCCGCAGGAGCAG 279  
DB 1 PhePheArgGluAenLeuAlaPheProGlnGlyAlaArgGluPheProSerSerGluGln 20  
QY 280 AACCGCGCAACAGCCCGCACAGCGCGGAGCTGCAGGTG---CGCGGCGCAACCCCGCGC 336  
DB 21 ThrArgAlaAenSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40  
QY 337 AGCGAGCGCGCGCGAGCGCGAGGC-----ACCCTGAACCTTCCCGCAGATCACCTG 390  
DB 41 SerGluThrGlyAlaGluArgGlnGlyLeValSerPheSerPheProGlnThrLeu 60  
QY 391 TGGCAGCGCCCGCTGGTGAGCATCAAGTGGCGCGCAGATCAAGGAGCGCCCTCTCTGAC 450  
DB 61 TrpGlnArgProValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAsp 80  
QY 451 ACCGGCGCGCAGCACCGCTGCTCGAGGAGATGAGCTGCCCGCAAGTGAAGCCCAAG 510  
DB 81 ThrGlyAlaAspAspThrValLeuGluGluLeuAenLeuProGlyLysTrpLysProLys 100  
QY 511 ATGATCGCGCGCATCGCGGCTTATCATAGTGGCGCAGTACGACGAGATCCTGATCGAG 570  
DB 101 MetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu 120  
QY 571 ATCTCGCGCAAGAGGCGCATCGGACCGCTGATCGCGCCCGCCCGCGTGAACATCATC 630  
DB 121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 140  
QY 631 GCGCGCAACATGTGACCCAGCTGGCTGACCTGACCTGAACTTCCCATCAGCCCATCGAG 690  
DB 141 GlyArgAenMetLeuThrGlnIleGlyCysThrLeuAenPheProIleSerProIleGlu 160  
QY 691 ACCGTGCGCTGAGCTGAAGCCCGCATGACCGCGCATGACCGCGCCCGCGAGTGAAGCTG 750  
DB 161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180  
QY 751 ACCGAGGAGAGATCAAGGCGCTGACCGCATCTGCGAGGAGATGGAGAGAGGCGCAAG 810  
DB 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200  
QY 811 ATCACCAGATCGCGCGCGAGAACCCCTTACAAACACCCCGTGTTCGCGCATCAAGAAGAAG 870  
DB 201 IleLeuLysIleGlyProGluAenProTrpAsnThrProValPheAlaIleLysLysLys 220  
QY 871 GACAGCACCAAGTGGCGCGAGCTGGTGGACTTCGCGAGCTGAACAGCGCGCACCGAGAC 930  
DB 221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAenLysArgThrGlnAsp 240  
QY 931 TTCTGGAGGTGCGAGCTGGGCGATCCCGCCCGCGCTGAGAGAGAGAGAGAGCGGTG 990  
DB 241 PheTrpGluValGlnLeuGlyIleProHlaProAlaGlyLeuLysLysLysSerVal 260  
QY 991 ACCGTGCTGAGCGTGGCGAGCGCTACTTTCAGCTGCGCCCTCGGACGAGCATTCGCGCAAG 1050  
DB 261 ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArgLys 280  
QY 1051 TACACCGCTTACCATCCCGCAGCATCAACAGAGACCCCGCGCATCCCGTACCATGATC 1110

DB 281 TyrThrAlaPheThrIleProSerIleAenAenGluThrProGlyIleArgTyrGlnTyr 300  
QY 1111 AACGTGCTGCCCCAGGCGTGAAGCGCAGCGCCCGCATCTTCCAGAGCAGCATGACCAAG 1170  
DB 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
QY 1171 ATCTGAGAGCCCTTCCGGCGCCCGCAACCCCGAGATCGTGATCTACCAG-----GCCGCC 1224  
DB 321 IleLeuGluProPheArgThrLysAenProGluIleValIleTyrGlnTyrMetAspAsp 340  
QY 1225 CTGTACGTGGCGCAGCGACCTGAGATCGCGCAGCAGCGCGCAAGATCGAGAGCTGCGC 1284  
DB 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuLeuArg 360  
QY 1285 AAGCACCTGCTGCTGGGCTTCCACCAACCCCGCAGCAAGAACCAAGAGAGAGCGCC 1344  
DB 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
QY 1345 TTCTGTGATGGGCTAGAGCTGACCCCGACAGTGGACCGTGCAGCCCATCGAGCTG 1404  
DB 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400  
QY 1405 CCGGAGAGGAGAGCTGACCGCTGACACATCCAGACATCCAGAGCTGGTGGCAAGCTGAACCTGG 1464  
DB 401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420  
QY 1465 GCCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTCTCGCGCGGCC 1524  
DB 421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440  
QY 1525 AAGCCCTGACCGACATCGTGCCTCGACCGAGGCGCGAGCTGGAGCTGGCGAGAAC 1584  
DB 441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuLeuAlaGluAsn 460  
QY 1585 CGCGAGATCTCGCGCAGCGCGCTGACGGCTGCTACGACCGCGCAGCGCAAGAGCTGCTG 1644  
DB 461 ArgGluIleLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
QY 1645 GCCGAGATCCAGAGAGGCGCGACGACGAGTGGACCTACCAAGATCTACGAGAGCCCTTC 1704  
DB 481 AlaGluIleGlnLysGlnGlyGlnTrpThrTrpThrTrpThrTrpThrTrpThrTrp 500  
QY 1705 AAGAACCTGAAGCCGCAAGTACGCCCAAGTGGCGACCGCGCCACACCAACGAGCTGAAG 1764  
DB 501 LysAenLeuLysThrGlyLysTrpAlaArgIleLysSerAlaHisThrAsnAspValLys 520  
QY 1765 CAGCTGACCGCGCGCTGACAGAGATCGCATCGAGAGCATCGTGATCTGGGGCAAGACC 1824  
DB 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
QY 1825 CCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884  
DB 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrTrpTrp 560  
QY 1885 CAGCGCACCTGTGATTCGCGAGTGGAGTTCTGTGAACACCCCGCTGGTGAAGCTGTGG 1944  
DB 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
QY 1945 TACCAGCTGAGAGAGGAGCCCATCATCGCGCGCAGACCTTCTACGTGGAGCGCGCGCC 2004  
DB 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
QY 2005 AACCGCAGAGCAAGATCGCGCAGCGCGCTGACGTGACCGACCGCGCGCGCGCAGAGATC 2064  
DB 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
QY 2065 GTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGAGCGCATCTCCAGCTGGCCCTG 2124  
DB 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
QY 2125 CAGGACAGCGCGCAGGAGTGAACATCGTACCGCAGCAGCAGTACCGCTGGGCGATCATC 2184  
DB 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660

Qy	2185	CAGGCCAGCCCGACAGAGCGAGCGAGCGAGCTGCTGTGAACACAGATCATCGAGCGAGCTGATC	2244
Db	661	GlnAlaGlnProAspLysSerGluSerGluValAsnGlnIleGluGlnLeuIle	680
Qy	2245	AAGAGGAGGAGGTGTACCTGAGCTGGGTGCCGCCCCACAGGGCATCGGGCGCAACGAG	2304
Db	681	GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyAsnGlu	700
Qy	2305	CAGATGCACAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCTTCGAGCGGCATCGAT	2364
Db	701	GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp	720
RESULT 12			
AAO30963	ID	AAO30963 standard; protein; 1003 AA.	
XX	AC	AAO30963;	
XX	DT	06-OCT-2003 (first entry)	
XX	DE	HIV pol protein.	
XX	KW	Human immunodeficiency virus; HIV; pol; IV9; immunotherapy; infection;	
XX	KW	epitope; vaccine; enzyme.	
XX	OS	Human immunodeficiency virus.	
XX	PN	WO2003051285-A2.	
XX	PD	26-JUN-2003.	
XX	PF	29-OCT-2002; 2002WO-US034620.	
XX	PR	29-OCT-2001; 2001US-0345116P.	
XX	PA	(GENZ ) GENZYME CORP.	
XX	PA	(GHO ) GEN HOSPITAL CORP DBA MASSACHUSETTS GEN.	
XX	PI	Nicolette CA, Walker BD;	
XX	DR	WPI; 2003-541603/51.	
XX	PT	Novel synthetic therapeutic peptide compounds useful for inducing an	
XX	PT	immune response in a mammal, specifically cross-react with HIV IV9	
XX	PT	epitopes.	
XX	PS	Claim 8; Page 65-67; 67pp; English.	
XX	CC	The invention relates to synthetic therapeutic peptide compounds that	
XX	CC	specifically cross-react with HIV (human immunodeficiency virus) pol	
XX	CC	protein IV9 epitopes. The invention is useful for treating HIV infection.	
XX	CC	The invention is also used in immunotherapy and as vaccines. The present	
XX	CC	sequence is HIV pol protein	
XX	SQ	Sequence 1003 AA;	
Alignment Scores:			
Pred. NO.:	9.96e-194	Length:	1003
Score:	3497.00	Matches:	655
Percent Similarity:	95.56%	Conservative:	34
Best Local Similarity:	90.85%	Mismatches:	26
Query Match:	76.47%	Indels:	6
DB:	6	Gaps:	3
US-09-610-313B-31 (1-2463) x AAO30963 (1-1003)			
Qy	220	TTCTTCGCGAGGACCTGGCTTCCCGAGGGCAAGGCCGCGAGTTCCTCCGAGCGAGCAG	279
Db	1	PhePheArgGluuAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln	20
Qy	280	AACCGCGCCAAAGCCCCACCGCGCGAGCTGCGAGTGTG-----CGCGCGCAACACCC	333

```
QY 1402 CTGCCCCGAGAGAGCTGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1461
DB 1403 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGGCCAGCAGACTACCCCGGATCAAGGTGGCGAGCTGTCAGAGCTGCTGGCGGC 1521
DB 1463 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 421 TrpAlaSerGlnIleTrpProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1522 GCCAAGCCCTGACCGACATCGTCCCTGACCCAGGAGGCGGAGCTGGAGCTGGCGAG 1581
DB 1523 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1582 AACCGCAGATCTCGCGAGCCCGTGCAGCGCTGTACTACGACCCCGAGCAAGACTG 1641
DB 1583 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1642 GTGGCCGAGATCCAGAGAGCGGCGACAGCAGTGGACCTACAGATCTACCGAGGCC 1701
DB 1643 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 481 IleAlaGluIleGlnLysGlnGlyGlnGlnTrpThrTyrGlnIleTyrGlnGluPro 500
QY 1702 TTCAAGAACCTGAGACCGGCAAGTACGCCAAGATGCCACCGCCACACCAACGAGCTG 1761
DB 1703 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCGAGCGCTGCAGAGATGCCATGAGAGCATCTGTGATCTGGGCAAG 1821
DB 1763 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 521 LysGlnLeuThrGluAlaValGlnLysIleThrGluSerIleValIleTrpGlyLys 540
QY 1822 ACCCCCAAGTCCGCTCCCTCCATCCAGAGAGACTGGGAGACTGGTGACCGACTAC 1881
DB 1823 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1882 TGGCAGCGCCACTGTGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGTGGAAGCTG 1941
DB 1883 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1942 TGTATCAGCTGGAGAGAGCCCATCATCGCGCCGAGACCTTCTACGTGGAGCGGCC 2001
DB 1943 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCAGAGCAAGATCGGAGCGCGCTACGTGACCGAGCCGCGCGGCGAGAG 2061
DB 2003 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2062 ATCGTAGCTGACCGAGACACCAACAGAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2121
DB 2063 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
QY 2122 CTGCAGGACAGCGCGAGGTGAACATCGTGACCGAGCCAGTACGCCCTGGGCATC 2181
DB 2123 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCAGCGCCAGCCGCAAGAGGAGAGAGCTGTGTGAACCGATCATCTGAGAGCTG 2241
DB 2183 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAAGAGAGGTGTACCTGAGCTGGTGCCTCCGCGCAGGCGCATCCGCGGCAAC 2301
DB 2243 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2302 GAGCAGATCGACAGCTGGTGAGCAAGGCGCATCCGAGGTGTGTCTCTGGAGCGGCATC 2361
DB 2303 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
DB 2363 : : : :
DB 721 Asp 721
```

RESULT 13

ADN36413

ID ADN36413 standard; protein; 1003 AA.

XX

```
AC ADN36413;
XX 15-JUL-2004 (first entry)
XX HIV protein for anti-HIV vaccine.
DE
XX anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX Human immunodeficiency virus 1.
OS
XX WO2004035006-A2.
PN
XX 29-APR-2004.
PD
XX 17-OCT-2003; 2003WO-US033112.
XX
XX 18-OCT-2002; 2002US-0419465P.
PR
XX (AARO-) AARON DIAMOND AIDS RES CENT.
XX
XX Huang Y, Ho DD, Chen Z;
XX WPI: 2004-348328/32.
XX N-PSDB; ADN36412.
XX
XX Nucleic acid vector comprising at least one HIV sequence operably linked
XX to a promoter and encoding a protein that does not assemble into viral
XX particles, useful in immunizing a subject against HIV infection.
XX
XX Disclosure; SEQ ID NO 27; 166pp; English.
XX
XX The invention relates to a nucleic acid vector comprising at least one
XX HIV sequence operably linked to a promoter and encoding a protein that
XX does not assemble into viral particles. The nucleic acid vector is useful
XX in immunizing a subject against HIV infection. This sequence corresponds
XX to a nucleic acid used in the invention.
XX
XX Sequence 1003 AA;
SQ
Alignment Scores:
Pred. No.: 9,96e-194 Length: 1003
Score: 3497.00 Matches: 655
Percent Similarity: 95.56% Conservative: 34
Best Local Similarity: 90.85% Mismatches: 26
Query Match: 76.47% Indels: 6
DB: Gaps: 3
US-09-610-313B-31 (1-2463) x ADN36413 (1-1003)
QY 220 TTCTTCGCGAGGACCTGGGCTTCCCGGCAAGGCCGCGAGTTCCCGAGGAGCAG 279
DB 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCAAACAGCCACCCAGCGCGAGCTGGAGGTG-----CGCGGCGCAACCCC 333
DB 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCGAGCGCGAGCGCGCGCTG-----AACTTCCCGCAGATCAC 387
DB 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
QY 388 CTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGCGCCAGATCAAGAGGAGCGCTGTG 447
DB 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGAGCAGACCGCTGTGCGAGAGATGAGCTGCCCGGCGAGTGAAGCCC 507
DB 81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGAGTACGAGCAGATCCTGATC 567
DB 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuLeu 120
```

```
QY 568 GAGATCTGCGGCAAGGCCATCGGCACCGTGTGATCGGCCCGCCACCCCGGTGAACATC 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGACCCAGCTGGGTGTGACCTGAACCTTCCCATCAGCCCGCATC 687
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTCCCGTGAAGCTGAAGCCCGCATCGAGCGGCCCGCCAGGTGAAGCAAGTGGCCC 747
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGAGAAATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAAAGAGGGC 807
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCCCGGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAAG 867
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCGGCGAGCTGAACGCGCACCCAG 927
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGGAGGTGAGCTGGGCATCCCGCACCCCGCGCTGGAAGAAAGAAAGAGC 987
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGACGTGGCGCAGCGCTACTTCAGCGTGCCCTGGAAGAGACTTCGCG 1047
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGCAGCATCAACACGAGACCCCGCATCCCGCTACCAG 1107
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGTGCCCCCAGGCTGGAAGGCGAGCCCGCAGCATCTTCACAGACGACGATGACC 1167
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCTACACAG-----GCC 1221
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340
QY 1222 CCCTGTACGTGGCGCAGCTGGAGATCGGCCAGCAGCCCGCCAGATCTCGAGGAGCTG 1281
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGTGCGCTGGCGCTTCACCAACCCCGCAGCAAGAACACCAAGAGAGGCC 1341
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysTrpThrValGlnProIleVal 380
QY 1342 CCCTTCTGTGGATGGCTACGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAG 1401
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCCCGAGAGAGAGCTGCAGCGCTGAACGACATCCAGAGCTGTGGCGCAGCTGAAC 1461
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGGCGCAGCAGATCTACCCCGCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGC 1521
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421 TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1522 GCCAAGGCCCTGACCGACATCGTGCCCTCACCAGAGAGCCCGAGCTGGAGCTGGCGGAG 1581
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
QY 1582 AACCGGAGATCTTCGCGGAGCCCGTGCACGGGTGTACTACGACCCCGCAGGAGCTG 1641
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480
QY 1642 GTGGCCGAGATCCAGAAGCAGGGGCCACGACGAGTGGACCTTACCAGATCTACCAGGAGCC 1701
```

```
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481 IleAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
QY 1702 TTCAAAGAACTGAGACCGGCAAGTACGCAAGATGGCAGCCGCCACACCAACGAGCTG 1761
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCCGAGCGCTGCAAGAGATCGCCATCGAGAGCATCGTGTGATCTGGGGCAAG 1821
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540
QY 1822 ACCCCCAAGTTCCGCTGCCATCCAGAGAGAGACCTGGGAGACCTGTGTGGACCGACTAC 1881
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1882 TGGCAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCCGCTGTGTGAAGCTG 1941
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580
QY 1942 TGGTACAGCTGGAGAAAGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGACGGCGCC 2001
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCAGACCAAGATTCGCAAGCGCGCTACGTGACCAACCGGGCGCGCAGAG 2061
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2062 ATCGTGAGCTGACCCAGACCAACCAACAGACGAGCTGAGCCGATCCAGCTGGCC 2121
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
QY 2122 CTCGACGACAGCGCGCAGGAGTGAACATCGTCAACGACAGCCAGTACGCGCTGGGCATC 2181
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCCAGCCCGACCGCCAGACGAGAGCGAGCTGTGTGAACCGAGATCATCGAGCAGCTG 2241
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAGGAGAGGTGTACTTGAGCTGGTGGCGCCGCCACCAAGGCGCATCGCGGCAAC 2301
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2302 GAGCAGATCGACAAGCTGTGTGAGCAAGGCGCATCCGCAAGGTGTCTTCTGGACGGCATC 2361
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db |||||
721 Asp 721
RESULT 14
AAP81861
ID AAP81861 standard; protein; 1002 AA.
XX
AC AAP81861;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 16-DEC-1990 (first entry)
XX
XX Sequence encoded by LAV MA L POL gene.
XX HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.
XX Lymphadenopathy-associated virus.
XX WO8707906-A.
XX
XX 30-DEC-1987.
XX
XX 22-JUN-1987; 87WO-EP000326.
XX
```

```

PR 23-JUN-1986; 86EP-00401380.
XX (INSP ) INST PASTEUR.
XX
XX Alison M, Sonigo P, Wainhobson S, Montagnier L, Alison M;
XX Sonigo P, Wainhobson S, Montagnier L;
XX
XX WPI; 1988-014396/02.
XX N-PSDB; AAN80437.
XX
XX New variants of lymphadenopathy associated virus (LAV) - used for prodn.
XX of DNA, antigens and antibodies used in diagnosis of AIDS and pre-AIDS.
XX
XX Claim 8; Fig 8A-8I; 72pp; English.
XX
XX LAV EL I (AAN80436) and LAV MA L (AAN80437) were isolated from the
XX peripheral blood lymphocytes of patients. Different AIDS virus isolates
XX concerned are designated by 3 letters of the patients name. Stable probes
XX including the DNA sequences can be used for detection of the new LAV
XX viruses or related viruses or DNA proviruses in eg. biological samples.
XX The proteins or peptides can be used for detection of antibodies induced
XX in vivo and present in biological fluids. The DNA can also be used for
XX the expression of LAV viral antigens for the prodn. of a vaccine against
XX LAV. The polypeptides can also be used for the prodn. of antibodies for
XX the detection of proteins related to the LAV viruses, partic. for
XX diagnosis of AIDS or pre-AIDS. (Updated on 25-MAR-2003 to correct PI
XX field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 1002 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1.39e-193 Length: 1002
XX Score: 3494.50 Matches: 652
XX Percent Similarity: 95.97% Conservative: 39
XX Best Local Similarity: 90.56% Mismatches: 24
XX Query Match: 76.42% Indels: 5
XX DB: 1 Gaps: 3
XX
XX US-09-610-313B-31 (1-2463) x AAP81861 (1-1002)
XX
XX QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCGCGAGTTCCTCCAGCGAGCAG 279
XX DB 1 PhePheArgGluAenLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20
XX
XX QY 280 AACCGCGCCCAACGCCCCAGCCAGCGCGAGCTGCGAGTG---CGCGCGCAACAACCCCGC 336
XX DB 21 ThrArgAlaenSerProThrSerArgGluLeuArgValTrpGlyGlyAsePheLysThrLeu 40
XX
XX QY 337 AGCGAGCGCGCGCGCGAGCGCGAGGCG-----ACCTGGAACCTTCCCGCGAGATCACCGTG 390
XX DB 41 SerGluThrGlyAlaGluArgGlnGlyLeValSerPheSerPheProGlnLeuThrLeu 60
XX
XX QY 391 TGGCAGCGCGCGCGCGAGCATCAAGGTGGCGCGCGAGATCAAGGAGGCGCGCTGCTGAC 450
XX DB 61 TrpGlnArgProValThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAseP 80
XX
XX QY 451 ACCGCGCGCCAGCACACCGTGTCTGCGAGGAGATGACCTGCGCGCAAGTGAAGCCCAAG 510
XX DB 81 ThrGlyAlaAsePheThrValLeuGluGluLeuAsePheProGlyLysTrpLysProLys 100
XX
XX QY 511 ATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGCGAGTACGACCGAGTCTGATCGAG 570
XX DB 101 MetileGlyLysLysAlaileGlyThrileLeuValGlyProThrProValAsePheLys 120
XX
XX QY 571 ATCTGCGCGAGAGGCGCATCGCGCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCATCATC 630
XX DB 121 ileCysGlyLysLysAlaileGlyThrileLeuValGlyProThrProValAsePheLys 140
XX
XX QY 631 GCGCGCAACATGCTGACCGCGTGTGACCGCGTGTGACCGCGTGTGACCGCGCGCGCGCGCG 690
XX DB 141 GlyArgAseMetLeuThrGlnileGlyCysThrLeuAsePheProLysProLysLeu 160
XX
XX QY 691 ACCGTGCGCGTGAAGCTGAAGCCCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 750
XX

```

---

```

Db 161 ThrValProValLysLeuLysProGlyMeAsePheProArgValLysGlnTrpProLeu 180
QY 751 ACCGAGGAGAGATCAAGGCGCTTCCCGCGAGGATCGGAGGAGGAGGAGGAGGAGGAGGAG 810
Db 181 ThrGluGluLysLysLysAlaLeuThrGluLysCysLysAsePheMetGluLysGluGlyLys 200
QY 811 ATCACCAAGATCGCGCGCGAGAACCCCTTACAAACACCCCGTGTTCGCGCATCAAGAGAGAG 870
Db 201 ileLeuLysLysLysGlyProGluAsePheProThrProValPheAlaileLysLysLys 220
QY 871 GACAGCACCAAGTGGCGCAAGCTGTGCGAGCTTCCCGCGAGCTGAACAGCGCACCCAGGAC 930
Db 221 AsePheThrLysTrpArgLysLeuValAsePheArgGluLeuAsePheArgGlnAseP 240
QY 931 TTCTGCGAGGTGCGAGTGGCGCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTG 990
Db 241 PheTrpGluValGlnLeuGlyLysProAlaileGlyLysLysLysLysLysLysLys 260
QY 991 ACCGTGCTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1050
Db 261 ThrValLeuAsePheValGlyAsePheLysPheSerValProLeuAsePheGluAsePhe 280
QY 1051 TACACCGCTTCCACCATCCCGCGCATCAACAGAGACCCCGCGCGCGCGCGCGCGCGCGTAC 1110
Db 281 TyrThrAlaPheThrLysProSerLysAsePheProGlyLysLysLysLysLysLys 300
QY 1111 AACGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1170
Db 301 AseValLeuProGlnGlyTrpLysGlySerProAlailePheGlnSerSerMetThrLys 320
QY 1171 ATCTGCGAGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1224
Db 321 ileLeuGluProPheArgThrLysAsePheProGluLysLysLysLysLysLysLys 340
QY 1225 CTGTACGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1284
Db 341 LeuTyrValGlySerAsePheLysLysLysLysLysLysLysLysLysLysLysLys 360
QY 1285 AAGCACCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1344
Db 361 GluHisLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 380
QY 1345 TTCTGTGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1404
Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAsePheTrpThrValGlnProLysLysLys 400
QY 1405 CCGGAGAGGAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1464
Db 401 ProAsePheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 420
QY 1465 GCGAGCGAGATCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1524
Db 421 AlaSerGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 440
QY 1525 AAGCGCTGACCGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1584
Db 441 LysAlaLeuThrAsePheValProLeuThrAlaGluLysLysLysLysLysLysLys 460
QY 1585 GCGGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1644
Db 461 ArgGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 480
QY 1645 GCGGAGATCTGCGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1704
Db 481 AlaGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 500
QY 1705 AAGAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1764
Db 501 LysAsePheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 520
QY 1765 CAGCTGACCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1824

```

Db 521 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 540  
 Qy 1825 CCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGGACCGACTACTGG 1884  
 Db 541 ProllysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpThrGluTrp 560  
 Qy 1885 CAGCGCACCTCGGATCCCGAGTGGAGTTCTGTAACACCCCGCTGTTGAAGCTGTGG 1944  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 580  
 Qy 1945 TACAGCTGAGAGAGGAGCCATCATCGCGCGGAGACCTTCTACGTGGACGGCGGCC 2004  
 Db 581 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 Qy 2005 AACCGCGAGACGATCGCAGAGCGCGTACGTGACACCGCGGCGCGCAGAGATC 2064  
 Db 601 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 Qy 2065 GTGAGCTGACCGAGACCAACCAAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTG 2124  
 Db 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 640  
 Qy 2125 CAGGACAGCGCGAGGAGTGAACTGTCGACGACGACGACGACGACGACGACGACGAC 2184  
 Db 641 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 Qy 2185 CAGCGCCAGCCGACAGAGGAGCGAGCTGCTGTAACACGACGATCATCGACGAGCTGATC 2244  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680  
 Qy 2245 AAGAAGGAGAAGGTGTACCTGAGCTGGGTGCGCGCCGACCAAGGGCATCGCGGCAACGAG 2304  
 Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlu 700  
 Qy 2305 CAGATCGACAAGCTGTGTGACAGGCGATCCGCAAGGTGCTGTTCTTGACGCGCATCGAT 2364  
 Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 15  
 ABR55489  
 ID ABR55489 standard; protein; 1003 AA.  
 AC ABR55489;  
 XX 11-AUG-2003 (first entry)  
 DT Amino acid sequence of a HIV pol protein.  
 XX Polymorphic marker; host response; HIV; AIDS related virus; vaccine;  
 KW hepatitis related virus; HCV; HBV; drug resistance; pol.  
 XX Human immunodeficiency virus.  
 OS WO2003035097-A1.  
 XX 01-MAY-2003.  
 PD 23-OCT-2002; 2002WO-AU001450.  
 XX 23-OCT-2001; 2001AU-00008425.  
 PR (BPIP-) EPIPOP PTY LTD.  
 PA Mallal S;  
 PI WPI; 2003-449231/42.  
 DR Determining the influence of variation in host genes on the selection of  
 PT microorganisms with protein substitutions, comprises typing individuals  
 PT of a cohort infected with a microorganism for an intrinsic polymorphic  
 PT marker.  
 XX Claim 22; Page 88-89; 157pp; English.

XX The specification describes a method of determining the influence of  
 CC variation in host genes on selection of microorganisms with protein  
 CC substitutions. The method comprises typing all individuals of a  
 CC population of patients infected with a microorganism for at least one  
 CC selected intrinsic polymorphic marker involved in the host response to  
 CC the presence of the microorganism. The method is useful for examining  
 CC selective pressures confronting a wide range of organisms that exhibit  
 CC pathogenic traits in a host, such as bacteria, fungi, mycobacterium,  
 CC viruses and virus-like particles; for examining microorganisms that have  
 CC adapted to evolve rapidly, including HIV and AIDS related viruses and the  
 CC hepatitis related viruses such as HCV and HBV. The method is useful for  
 CC designing a vaccine to prevent or delay the emergence of drug resistance  
 CC in patients treated with a particular drug specific for a microorganism  
 CC where the drug affects the replication of the microorganism at the  
 CC nucleotide or amino acid level. The present sequence represents a HIV pol  
 CC protein, which is expected to provide optimal cytotoxic T lymphocyte  
 CC (CTL) induced therapeutic protection to the cohort examined in that study  
 XX Sequence 1003 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1,48e-193 Length: 1003  
 Score: 3494.00 Matches: 654  
 Percent Similarity: 95.70% Conservative: 36  
 Best Local Similarity: 90.71% Mismatches: 25  
 Query Match: 76.40% Indels: 6  
 DB: 6 Gaps: 3  
 US-09-610-313B-31 (1-2463) x ABR55489 (1-1003)  
 Qy 220 TTCTTCCGCGAGACCTGGCTTCCCGGCAAGGCGCGCGGAGTTCGCCAGCGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheSerGluGln 20  
 Qy 280 AACCGCGCAACAGCGCCACCGACCGCGAGCTGCGGTGGCGGCGAC-----AACCCC 333  
 Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyLysAsnAsnSer 40  
 Qy 334 CGCAGCGAGCG 387  
 Db 41 ThrSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
 Qy 388 CTGTGCGAGCG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeu 80  
 Qy 448 GACACCG 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
 Qy 508 AAGATGATCG 567  
 Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleIleIle 120  
 Qy 568 GAGATCTGCGGCAAGAGCG 627  
 Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
 Qy 628 ATCGCGCGCAACATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687  
 Db 141 IleGlyArgAsnLeuLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIle 160  
 Qy 688 GAGACCGTCCCGTGAAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 Qy 748 CTGACCGAGGAGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 Qy 808 AAGATCACAAGATCG 867



Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
Qy 868 AAGGACACCAAGTGGCGCAAGCTGGTGAAGTTCGCGAGCTGAACAAGCGCAGCCAG 927  
Db 221 LysAspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
Qy 928 GACTTCTGGAGGTGACGTGGCATCCCCACCCCGCGCGCTGAAGAAGAAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLysLysSer 260  
Qy 988 GTGACCGTGTGGAGCTGGCGGCGCTTACTTTCAGCGTCCCTGGAGGAGCTTCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
Qy 1048 AAGTACACCGCTTACCACATCCCGAGCATCAACACGAGACCCCGCGCATCCGTACCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
Qy 1108 TACAACGTGTGCGCGAGGCTGAAGGGAGCGCCAGCATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
Qy 1168 AAGATCTCTGGAGCTTCCGCGCGCGCAACCCCGAGATCGTGATCTACCAG- ----GCC 1221  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
Qy 1222 CCCCTGTACGTGGCGAGCGACTCGAGATCGCGCAGCACCGCGCAAGATCGAGGAGCTG 1281  
Db 341 AspleuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
Qy 1282 CGCAAGCACCTGTGCGTGGGCTTCCACCAACCCCGACAGAAGACCGAGAGGAGCCC 1341  
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrTrpProAspLysLysHisGlnLysGluPro 380  
Qy 1342 CCCTTCTGTGGTGGCTACGAGCTGCACCCCGACAGTGGACCGTGCAGCCCATCGAG 1401  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
Qy 1402 CTGCCCGAGAGGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1461  
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
Qy 1462 TGGCCACGACAGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTGCTCGCGGC 1521  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
Qy 1522 GCCAAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCCGAG 1581  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
Qy 1582 AACCGCGAGATCCCTGGCGAGCCGTGCACGGCGTGTACTACGACCCCGACGAGGAGCTG 1641  
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480  
Qy 1642 GTGGCCGAGATCCAGAGAGCGGCCACGACGAGTGGACCTTACCAAGATCTACGAGGAGCCC 1701  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTyrTrpThrGlnIleTyrGlnGluPro 500  
Qy 1702 TTCAGAACCTGAAGACCGGCAAGTAGCCAGATGGCAGCGCCGACACCAAGAGAGCTG 1761  
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
Qy 1762 AAGCAGCTGACGAGGCGGTGCAGAAGATCGCATCGAGAGCATCGTGATCTGGGGCAAG 1821  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
Qy 1822 ACCCCCAAGTTCGCCCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC 1881  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyr 560  
Qy 1882 TGGCAGCCGACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

Qy 1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGCGCCCGAGACCTTTCTACGTGACGCGCC 2001  
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
Qy 2002 GCCAACCCGAGAGACCAAGATCGGCAAGCCCGGCTACGTGACCGACCGCGCGCGCAGAG 2061  
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620  
Qy 2062 ATCTGTAGCTGACCGAGACCAACCAAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2121  
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
Qy 2122 CTGCAGACGACCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCTCGGCATC 2181  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
Qy 2182 ATCCAGGCCCGCCGACGAGCGAGCGAGCGAGCTGTGAACACGATCATCGACGCTG 2241  
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
Qy 2242 ATCAAGAAGGAGAGGTGTACTCTGAGCTGGGTGCCGCCCAAGGGCATCGCGGCAAC 2301  
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700  
Qy 2302 GAGCAGATCGACAGCTGGTGCAGCAAGGCGCATCCGCAAGGTGCTTCTCTGACGCGCATC 2361  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
Qy 2362 GAT 2364  
Db 721 Asp 721

Search completed: June 2, 2005, 04:42:47  
Job time : 231.833 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 02:53:39 ; Search time 40.8333 Seconds  
(without alignments)

11607.276 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 4573

Sequence: 1 gtcgacgccaccatggccga.....gggtagcaccggtagaattc 2463

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-Q=/cgn2\_1/USPFO\_spool/US09610313/runat\_31052005\_155137\_15159/app\_query.fasta\_1.7893  
-DB=PIR\_79 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09610313 @CGN 1 1 189 @runat\_31052005\_155137\_15159 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3505.5	76.7	1002	2	S54378
2	3497.5	76.5	1002	1	GNLJND
3	3492	76.4	1003	1	GNVWIV
4	3475	76.0	1015	1	GNVWH3
5	3453	75.5	1003	2	T09440
6	3450	75.4	1003	1	B44001
7	3449	75.4	1003	1	GNVW22
8	3439	75.2	1012	1	GNVWYL
9	3211	70.2	1027	1	GNLJ51
10	3100	67.8	912	2	S33980
11	3094	67.7	902	2	T01668
12	2771	60.6	559	2	B47175
13	2751	60.2	559	2	A47175
14	2350	51.4	1039	2	S46347

15	2209	48.3	1055	2	S53092	pol polyprotein -
16	2205.5	48.2	1055	1	GNLJST	HIV-1 retropepsin
17	2196.5	48.0	1056	1	GNLJG3	HIV-1 retropepsin
18	2196	48.0	1036	1	GNLJG2	HIV-1 retropepsin
19	2188	47.8	1034	1	GNLJCA	HIV-1 retropepsin
20	2187.5	47.8	1035	1	GNLJG3	HIV-1 retropepsin
21	2182.5	47.7	1019	2	T11560	pol polyprotein -
22	2168.5	47.4	1009	2	S28081	pol polyprotein -
23	2150.5	47.0	1032	2	S12153	pol polyprotein -
24	2073	45.3	1058	2	S08436	pol polyprotein -
25	2053	44.9	1061	1	GNLJG4	HIV-1 retropepsin
26	2019.5	44.2	1054	1	GNLJG5	HIV-1 retropepsin
27	1963.5	42.9	656	2	S30484	pol polyprotein -
28	1934.5	42.3	656	2	S30483	pol polyprotein -
29	1390	30.4	1124	1	S23820	pol polyprotein -
30	1387	30.3	1124	1	GNLJFP	HIV-1 retropepsin
31	1344	29.4	1124	2	B45557	HIV-1 retropepsin
32	1148.5	25.1	1109	1	B45345	HIV-1 retropepsin
33	1134	24.8	1146	1	GNLJ22	HIV-1 retropepsin
34	1134	24.8	1146	1	GNLJEW	HIV-1 retropepsin
35	1131.5	24.7	1087	2	JQ1162	Pol protein - Maed
36	1128.5	24.7	1086	1	B46335	HIV-1 retropepsin
37	1121	24.5	1101	1	B45390	HIV-1 retropepsin
38	1117	24.4	1101	1	GNLJVS	HIV-1 retropepsin
39	1114.5	24.4	1145	1	GNLJEV	HIV-1 retropepsin
40	1090	23.8	216	2	S43127	reverse transcript
41	1071	23.4	219	2	S32152	RNA-directed DNA p
42	1067	23.3	219	2	S32118	RNA-directed DNA p
43	1067	23.3	219	2	S32139	RNA-directed DNA p
44	1065	23.3	219	2	S32089	RNA-directed DNA p
45	1065	23.3	219	2	S32160	RNA-directed DNA p

#### ALIGNMENTS

##### RESULT 1

S54378

pol polyprotein - human immunodeficiency virus type 1 (fragment)

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

R;Accession: S54378

R;Theodore, T.; Buckler-White, A.J.

submitted to the EMBL Data Library, July 1989

A;Reference number: S54377

A;Accession: S54378

A;Status: preliminary

A;Molecule type: genomic RNA

A;Residues: 1-1002 <THE>

A;Cross-references: UNIPROT:P12499; EMBL:M22639; NID:G329377; PIDN:AAA45366.1; PID:G3293

C;Superfamily: pol polyprotein

C;Keywords: polyprotein

##### Alignment Scores:

Pred. No.:	2.59e-155	Length:	1002
Score:	3505.50	Matches:	655
Percent Similarity:	95.69%	Conservative:	34
Best Local Similarity:	90.97%	Mismatches:	26
Query Match:	76.66%	Indels:	5
DB:	2	Gaps:	3

US-09-610-313B-31 (1-2463) x S54378 (1-1002)

Qy	220	TTCTTCGCGGAGGACCTGGCTTCCCGGAGGAGCCGCGAGTTCCCGAGCGAGCAG	279
Db	1	PhapheargluaspLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerSerGluGln	20
Qy	280	AACCGCGCCCAACGCCCGGAGCGAGCTGCGAGTGGCGGC---GACAAACCCCGC	336
Db	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu	40
Qy	337	AGCGAGCGCGCGCGAGCGGACCCCTGACCTTC-----CCCCAGATCACCCCTG	390
Db	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu	60

QY	391	TGCAGCGCCCTGTGTGAGCATCAAGTGGGGCCGAGATCAAGAGGCGCTCTGCTGGAC	450
Db	61	TrpGlnArgProLeuValThrIleLysAlaLeuLysGluAlaLeuLeuAsp	80
QY	451	ACGGCGCCGACGACACACCCTGTGTGGAGGAGATGAGCCTGCCCGCGAAGTGGAGCCCAAG	510
Db	81	ThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys	100
QY	511	ATGATCGCGCGCATCGCGGCTTCATCAAGTGGCCAGTACCACAGATCCTGTGATCGAG	570
Db	101	MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu	120
QY	571	ATTCGCGCAAGAGGCATCGGCACCGTCTGATCGGCCCCACCCCGCTGAACATCATC	630
Db	121	IleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle	140
QY	631	GGCCGCAACATGTGACCCAGCTGGGTGCACCTGAACTTCCCATCAGCCCATCGAG	690
Db	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
QY	691	ACGTGCCCCGTGAAGTGAACCCCGGCATCGACGGCCCCAAGTGAACGAGTGGCCCCCTG	750
Db	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
QY	751	ACCGAGGAGAAGATCAAGGGCCCTGACCGCCATCTGCGAGGAGATGGAGAAGGAGGCGCAAG	810
Db	181	ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys	200
QY	811	ATCACCAAGATCGGCCCGCCGAGAACCCCTACAACACCCCGCTGTTCGCCATCAAGAAGAAG	870
Db	201	IleSerArgValGlyProGluAsnProTrpAsnThrProIlePheAlaIleLysLysLys	220
QY	871	GACAGCACCAAGTGGCGCAAGCTGTGTGGACTTCGGCGAGCTGAAACAAGCGCACCCAGGAC	930
Db	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
QY	931	TTCTGGAGAGTGCAGCTGGGCATCCCCACCCCGCGGCTGGAAGAAGAAGAAGAGCGTG	990
Db	241	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerVal	260
QY	991	ACCGTGTGCAGCTGGCGCAGCGCTACTTCAGCGTGCCTTCGACGAGGACTTTCGCGAAG	1050
Db	261	ThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspLysAspPheArgLys	280
QY	1051	TACACCGCCTTACCATTCCCAGCATCAACAAGAGACCCCGGCATCCGCTACCAAGTAC	1110
Db	281	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGlnTrp	300
QY	1111	AACGTGCTGCCCGAGGCTCGAAGGGCAGCCCGAGCATCTTCAGAGCAGCATGACCAAG	1170
Db	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
QY	1171	ATCTGTGAGCGCTTCCGCGCCGCAACCCCGAGATCTGTGATCTACAG-----GCCCC	1224
Db	321	IleLeuGluProPheArgLysGlnAsnProGluIleValIleTrpGlnTrpMetAspAsp	340
QY	1225	CTGTACTGTGGCAGCGACCTCGAGATCGGCCAGCACCGCGCCCAAGATCGAGGAGCTGCGC	1284
Db	341	LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	360
QY	1285	AAGCACCTGTGGCTGGCTGGCTTCCACACCCCGACAGAAGACCAAGAGGAGGCCCC	1344
Db	361	GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	380
QY	1345	TTCTGTGGATGGGCTACGAGCTGACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTG	1404
Db	381	PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu	400
QY	1405	CCCGAGAAGGAGAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGCGCAAGCTGAAC	1464
Db	401	ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	420

RESULT 2

## RESULTS

GNTJND  
HIV-1 reverse transcriptase (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK1)

N; Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA

C;Species: human immunodeficiency virus type 1, HIV-1

A;Note: host Homo sapiens (man)

C:\Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text\_change 09-Jul-2004

C>Date: 30 Dec 1971 #BQQWQCC\_10-18-68  
C:Accession: J00067

-

C/ACCESSION: UQ0067  
P.SNIRE R : sire J : Zachar. V : Rev. F. : Barre-Sinoussi. F. : Galibert. F. : Hampe,

R;Spire, B.; Spire, J.; Zacher, F.; Garibelli, F.; Garibelli, F.; Raggio  
Congo 01 275-284 1989

Gene 81, 275-284, 1989

A;Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus type 1

A:Molecule type: DNA  
A:Residues: 1-1002 <SPI>  
A:Cross-references: UNIPROT:P18802; GB:M27323; NID:g328154; PIDN:AAA44869.1; PID:g328158  
F:56-154/Product: retropesin #status predicted <RTP>  
C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
C:Genetics:  
A:Gene: pol  
C:Superfamily: pol polypeptide  
C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleoc  
F:56-154/Product: retropesin #status predicted <RTP>  
F:80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
Pred. No.: 6.08e-155 Length: 1002  
Score: 3497.50 Matches: 655  
Percent Similarity: 95.69% Conservative: 34  
Best Local Similarity: 90.97% Mismatches: 26  
Query Match: 76.48% Indels: 5  
DB: 1 Gaps: 3

US-09-610-313B-31 (1-2463) x GNLJND (1-1002)

QY	220	TTCTTCGCGAGGACCTGGCTTCCCGCAGGGCAAGCCGCGAGTTCCCGCAGCGAGCAG	279
DB	1	PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluPheSerSerGluGln	20
QY	280	AACCGCGCCAAACAGCCCGCCAGCGCGAGCTGCAGTG---CGCGCGCAACCCCGCGC	336
DB	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu	40
QY	337	AGCGAGCGCGCGCGAGCGCGAGCGCGAGCGCGCGCTG-----AACTTCCCGCGAGTCA	390
DB	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnLeuThrLeu	60
QY	391	TGGCAGCGCGCGCGCGAGCGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG	450
DB	61	TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp	80
QY	451	ACCGCGCGCGCGCGCGAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	510
DB	81	ThrGlyAlaAspThrValLeuGluLeuAsnLeuProGlyLysTrpLysProLys	100
QY	511	ATGATCG	570
DB	101	MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGlu	120
QY	571	ATCTGCGCGCAAGAGCG	630
DB	121	IleCysGlyTyrlsAlaMetGlyThrValLeuValGlyProThrProValAsnIle	140
QY	631	GGCGCGCAACATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	690
DB	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
QY	691	ACCGTCCCGTGAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	750
DB	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
QY	751	ACCGAGGAGAAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	810
DB	181	ThrGluGlnIleLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys	200
QY	811	ATCACCAAGATCG	870
DB	201	IleSerArgIleGlyProGluAsnProTyrlsThrProIlePheAlaIleLysLysLys	220
QY	871	GACAGCAACCAAGTGGCGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG	930
DB	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
QY	931	TTCTGCGAGGTGCGAGTGGCGATCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	990
DB	241	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysLysLys	260

QY	991	ACCGTGTGAGCGTGGCGGAGCGCTACTTTCAGCGTGGCGCGCGCGCGCGCGCGCGCG	1050
DB	261	ThrValLeuAspValGlyAspAlaTyrlsPheSerValProLeuAspGluAspPheArgLys	280
QY	1051	TACACGCGCTTACCACCTCCCGCAGCATCAACACAGAGACCCCGCGCATCCGCTACCA	1110
DB	281	TyrlsAlaPheThrIleProSerIleAsnAsnGluThrProGlyLysIleArgTyrlsGln	300
QY	1111	AACGTGTGCTGCCCGCAGGCTGGAAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG	1170
DB	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
QY	1171	ATCTGCGAGCGCTTCCCG	1230
DB	321	IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrlsGlnTyrlsMetAsp	340
QY	1225	CTGTACGTGGCGCAGCGACCTGGAGATCGGCGAGCGCGCGCGCGCGCGCGCGCGCGCG	1284
DB	341	LeuTyrlsValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	360
QY	1285	AAGCACCTGTGCTGGCGCTGGCGCTTCCACACCGCGCGCGCGCGCGCGCGCGCGCGCG	1344
DB	361	GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	380
QY	1345	TTCTGTGTGATGGCTACGAGCTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1404
DB	381	PheLeuTrpMetGlyTyrlsGluLeuHisProAspLysTrpValGlnProIleAsnLeu	400
QY	1405	CCCGAAGAGAGAGCTGGACCGGTGACGACATCCAGAGCTGTGCGCGCGCGCGCGCGCG	1464
DB	401	ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	420
QY	1465	GCACGCGAGATCTACCG	1524
DB	421	AlaSerGlnIleTyrlsAlaGlyIleLysValLysGlnLeuCysLysLeuLeuAspGlyThr	440
QY	1525	AAGCGCTGACCGACATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1584
DB	441	LysAlaLeuThrGluValProLeuThrGluAlaGluLeuGluLeuAlaGluAsn	460
QY	1585	CGCGAGATCTGCG	1644
DB	461	ArgGluIleLysLysGluProValHisGlyValTyrlsAspProSerLysAspLeuIle	480
QY	1645	GCAGAGATCCAGAGAGCG	1704
DB	481	AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrlsGlnIleTyrlsGlnProPhe	500
QY	1705	AAGACCTGAAGACCGCGCAAGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1764
DB	501	LysAsnLeuLysThrGlyLysTyrlsAlaArgThrArgGlyAlaHisThrAsnAspValLys	520
QY	1765	CAGCTGACCGAGCGCGTGCAGAGATCGCCATGCGAGAGCATCGTATCTCGCGCGCGCGCG	1824
DB	521	GlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr	540
QY	1825	CCCAAGTTCGCGCTGCCATCCAGAGAGGAGACCTGGGAGACCTGGTGGCGCGCGCGCG	1884
DB	541	ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrlsTrp	560
QY	1885	CAGCGCACCTGGATCCCGCGAGTGGAGTTCGTGAACACCGCGCGCGCGCGCGCGCGCGCG	1944
DB	561	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp	580
QY	1945	TACAGCTGAGAGAGGAGCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2004
DB	581	TyrlsGlnLeuGlnLysGluProIleGlyAlaGluThrPheTyrlsValAspGlyAlaAla	600
QY	2005	AACCGCGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGCGCGCGCGCGCGCGCGCGCG	2064
DB	601	AsnArgGluThrLysLeuGlyLysAlaGlyTyrlsValThrAspArgGlyArgGlnLysVal	620
QY	2065	GTGAGCGCTGACCGAGAGACCAACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	2124

Db	621	ValPropheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu	640	
QY	2125	CAGGACGCGGAGCGAGGTGAACATCGTACCGACGACGAGCGAGTACGCGCTGGGCATCATC	2184	
Db	641	GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTrpAlaLeuGlyIleIle	660	
QY	2185	CAGGCCAGCCGACAGAGCGAGAGCGAGCTGGTGAACAGATCATCAGAGAGCTGATC	2244	
Db	661	GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle	680	
QY	2245	AAGAAGAGAGGTGTACCTGAGCTGGTGGTCCCGCCACAAAGGCGATCGCGCGCAACGAG	2304	
Db	681	LysLysGlnLysValTrpLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsnGlu	700	
QY	2305	CAGATCGACAGCTGTGTGACGAGGCGATCCGCAAGGTGTGTCTCGAGCGCATCGAT	2364	
Db	701	GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp	720	
RESULT 3				
GNVWL				
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LAV-1a)				
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly				
C;Species: human immunodeficiency virus type 1, HIV-1				
A;Note: host Homo sapiens (man)				
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002				
C;Accession: A03966				
R;Main-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.				
Cell 40, 9-17, 1985				
A;Title: Nucleotide sequence of the AIDS virus, LAV.				
A;Reference number: A90866; MUID:85099333; PMID:2981635				
A;Accession: A03966				
A;Molecule type: DNA				
A;Residues: 1-1003 <WAI>				
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re				
C;Genetics:				
A;Gene: pol				
C;Superfamily: pol polyprotein				
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle				
F;57-135/Product: retropepsin #status predicted <RP>				
F;81/Active site: Asp (shared with dimeric partner) #status experimental				
Alignment Scores:				
Pred. No.:	1,09e-154	Length:	1003	
Score:	3492.00	Matches:	655	
Percent Similarity:	95.42%	Conservative:	33	
Best Local Similarity:	90.85%	Mismatches:	27	
Query Match:	76.36%	Indels:	6	
DB:	1	Gaps:	3	
US-09-610-313B-31 (1-2463) x GNVWL (1-1003)				
QY	220	TTCTTCGGGAGGACTGGCTTCCCGGAGGCGCGGAGTTCCTCCCGAGGAGCG	279	
Db	1	PhePheArgGluAspLeuAlaPheLeuGlnGlyLeuAlaArgGluPheSerSerGluGln	20	
QY	280	AACCGCGCAACGCCCGGAGCGGAGCTCGAGTG-----CGCGCGCAACCC	333	
Db	21	ThrArgAlaAsnSerProThrArgGluGlnValTrpGlyArgAspAsnSer	40	
QY	334	CGCAGCGAGCGCGCGGAGCGCGGAGCGGAGCGCGCTG-----AACTTCCCGGAGATCAC	387	
Db	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr	60	
QY	388	CTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGGCGGCGAGTCAAGGAGCGCTGTG	447	
Db	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu	80	
QY	448	GACACGGCGCGGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG	507	
Db	81	AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro	100	
QY	508	AAGATGATCGGCGGCGGCTTCATCAAGGTGGCGGAGTACGACCATCTCTGATC	567	

Db	101	LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIle	120	
QY	568	GAGATCTGGGCAAGAGGCGCATCGGACCGGTCTGATCGGCGCCACCCCGTGACATC	627	
Db	121	GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140	
QY	628	ATCGGCGCAACATGCTGACCCAGCTGGGCTGCACCTCAACTTCCCATCAGCCCCATC	687	
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160	
QY	688	GAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGCCCAAGGTGAAGCGATGGCCC	747	
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180	
QY	748	CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAGGAGGCG	807	
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200	
QY	808	AAATCACCAAGATCGGCGCGGAGAACCCCTACACACCCCGTGTTCGCCCATCAAGAAG	867	
Db	201	LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys	220	
QY	868	AAGGACAGCACCAAGTGGCGCAAGCTGTGGACTTCCCGGAGCTGAACAGCGCACCCAG	927	
Db	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240	
QY	928	GACTTCTGGAGGTGCGCTGGCGCATCCCGCCCGCGCGCTGAGAGAGAGAGAGAGC	987	
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260	
QY	988	GTGACCGTCTGACGCTGGCGGCGAGCTACTTTCAGGTGCGCTGAGCAGGACTTCCGC	1047	
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg	280	
QY	1048	AAGTACACCGCTTCCATCCCGAGCATCAACAGAGACCCCGCGCATCCGCTACCAG	1107	
Db	281	LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln	300	
QY	1108	TACACGTCTGCGCGGCGGCGGAGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG	1167	
Db	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320	
QY	1168	AAGATCTCGAGCGCTTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1221	
Db	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	340	
QY	1222	CCCCTGTAGTGGGCGAGCGACCTGGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGG	1281	
Db	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360	
QY	1282	CGCAGCACCTGCTGCGTGGGCGGCTTCCACCCCGGCGGCGGCGGCGGCGGCGGCGG	1341	
Db	361	ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro	380	
QY	1342	CCCTTCTGTGATGGGCTACGAGCTGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGG	1401	
Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	400	
QY	1402	CTGCGCGAGAGAGAGTGGACCGGTGAACGACATCCAGAGCTGTGGCGCAAGCTGAAC	1461	
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420	
QY	1462	TGGGCGGAGCAGATCTACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1521	
Db	421	TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly	440	
QY	1522	GCCAGGCGCTGACCGACATCGTGCCTCCCTGACCGGAGGCGGCGGCGGCGGCGGCGG	1581	
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuGluLeuAlaGlu	460	
QY	1582	AACCGGAGATCTCTGCGGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	1641	





Db	281	AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACAACAGAGACCCCGGCATCCGCTACCAGTACAACGTGCTGCCCCAGCGCTGG	1131
Db	301	SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp	320
Qy	1132	AAGGGCAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCC	1191
Db	321	LysGlySerProAlaIlePheGlnSerMetThrLysIleLeuGluProPheLysLys	340
Qy	1192	CGCAACCCCGAGATCGTGATCTACGAG-----GCCCCCTGTACGTGGGAGCGACCTG	1245
Db	341	GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu	360
Qy	1246	GAGATCGGCAGCACCGCGCCAGCATCGAGAGCTGCGCAAGCACCTGCTGCGTGGCGG	1305
Db	361	GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly	380
Qy	1306	TTCACCAACCCCGACAAGAACACACGAGGAGCGCCCTTCTCTGTGGATGGGTACGAG	1365
Db	381	LeuThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrGlu	400
Qy	1366	CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGA	1425
Db	401	LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr	420
Qy	1426	GTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACCTGGGCGACCCAGATCTACCCG	1485
Db	421	ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly	440
Qy	1486	ATCAAGGTGCGCAGCTGTGCAAGCTGTGCGCGCGCCAAAGGCCCTGACCGACATCGT	1545
Db	441	IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle	460
Qy	1546	CCCTGTACCGAGAGCGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGGAGCCC	1605
Db	461	ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluPro	480
Qy	1606	GTGCACCGCGTGACTACGACCCCGACGACCTGTCGCGAGATCCAGAGCAGCGGC	1665
Db	481	ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly	500
Qy	1666	CACGACGAGTGGACCTACCAAGATCTACGAGAGCCCTTCAAGAACCTGAAGACCGGCA	1725
Db	501	GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys	520
Qy	1726	TACGCCAAGATGCGCACCGGCCACACCAACGACTGAAGCAGCTGACCGAGGCGGTGAG	1785
Db	521	TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln	540
Qy	1786	AAGATCGCCATGGAGACATCGTATCTGGGCGAGACCCCAAGTTCGCCCTGCCCATC	1845
Db	541	LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle	560
Qy	1846	CAGAAGGAGACTCGGAGACCTGTGACCACTACTGGCAGGCGCACCTGGATCCCCGAG	1905
Db	561	GlnLysGluThrTrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu	580
Qy	1906	TGGAGTTCTGTGAACACCCCCCTGGTGAAGCTGTGGTACCAAGCTGGAGAGAGAGCCC	1965
Db	581	TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuLysGluPro	600
Qy	1966	ATCATCGCGCGGAGACCTTCTAGTGGACGGCGCGCCCAACCGGAGACCAAGATCGG	2025
Db	601	IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly	620
Qy	2026	AAGCGCGCTACGTGACCGACCGCGCGCGGAGAGATCGTGAAGCTGACCGAGACCA	2085
Db	621	LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr	640
Qy	2086	AACGACGAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCGAGGACAGCGGACGAG	2145

Db 641 AsnGlnIysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660

Qy 2146 AACATCGTGACCCAGACAGCAGTAGTACGCCCTGGGCATCATCCAGGCCAGCCGCAAGAGC 2205

Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleileGlnAlaGlnProAspLysSer 680

Qy 2206 GAGAGCGAGCTGTTGAACGAGATCATCGAGCAGCTCATCAAGAGGAGAAGGTGTACCTG 2265

Db 681 GluSerGluLeuValAsnGlnIleileGluGlnLeuIleLysGlnLysValTyrLeu 700

Qy 2266 AGCTGGTGTCGCCGCCCAAGGCGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGC 2325

Db 701 AlaIrrValProAlaHisGlyIleileGlyAsnGlnGlnValAspLysLeuValSer 720

Qy 2326 AAGGGCATCGCAAGGTGCTGTTCTCTCGAGCGGCATCGAT 2364

Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 5

T09440

pol polyprotein - human immunodeficiency virus type 1 (strain JRFL) (fragment)

C:Species: human immunodeficiency virus type 1, HIV-1

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C:Accession: T09440

R:Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; submitted to the EMBL Data Library, July 1996

A:Reference number: Z16673

A:Accession: T09440

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1003 <PAN>

A:Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779

C:Genetics:

A:Gene: pol

C:Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	7.03e-153	Length:	1003
Score:	3453.00	Matches:	646
Percent Similarity:	95.01%	Conservative:	39
Best Local Similarity:	89.60%	Mismatches:	30
Query Match:	75.51%	Indels:	6
DB:	2	Gaps:	3

US-09-610-313B-31 (1-2463) x T09440 (1-1003)

Qy 220 TTCTTCGGCAGAGCACTGGCTTCCCGCGGCAAGGCCGCGAGTTCGCCAGCGAGCAG 279

Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

Qy 280 AACCGCGCCAAAGCCGCCACCGCGCGAGCTCGAGTG-----CGCGCGCAACACCCC 333

Db 21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTyrpLysArgAspSerAsnSer 40

Qy 334 CGCAGCGAGCGCCGCGCCAGCGCCAGCGGCG-----ACCTGAACTTCCCGCCAGATCACC 387

Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60

Qy 388 CTGTGGCAGCGCCCTGTGTGACATCAAGTGGCGGCCAGATCAAGGAGGCCCTGCTCG 447

Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyGlnLeuLysGluAlaLeuLeu 80

Qy 448 GACACGGCGCCGACGACACCGTGTGGAGGAGATGAGCTGCGCGGCAAGTGGGAAGCCC 507

Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTyrLysPro 100

Qy 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGATGACGACGAGATCTTGATC 567

Db 101 LysMetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120

Qy 568 GAGATCTCGCGCAAGAGGCCATCGCGCGTCTGATCGCGCCCGCCACCGCCGTCACATC 627

Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

```
QY 628 ATCGGCGCCAAATGCTGACCCAGCTGGGCTGCACCTTGAATCTTCCCATCAGCCCCATC 687
Db 141 IleglyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGGCCGCTGACGTCAAGCCCGCATGACCGCCCAAGGTCAGACCTGGCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCCGAGGAGAAGATCAAGCCCTGACCCCATCTGCGAGGAGATGGAGAAGAGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCAACCAAGATCGGCCCGCCAGAACCCCTACAAACCCCGTGTTCGCCCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGCAGCAGCAAGTGGCGCAAGCTGGTGACTTCGCGAGCTGACACGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysLysThrGln 240
QY 928 GACTTCTGGGAGTGCAGCTGGCATCCCCACCCCGCCGCTGAAAGAAAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260
QY 988 GTGACCGTGTGGACGTGGCGCGCCCTTACTTACGCGTGCCTCGACGAGACTTCGCG 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGGCTTACCATCCCGATCCAGCATCAACACGAGACCCCGGATCCGCTACCAG 1107
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGTGCCCCAGGGCTGGAAGGGCAGCCCGACATCTTCCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCAAG-----GCC 1221
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
QY 1222 CCCCTGTACGTGGCGAGGACCTGGAGTCCGCGCAGACCCCGCATGAGTGCAGGAGCTG 1281
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGCTGCTGGGGCTTCACACCCCGCACAGAGAGACCAAGAGAGAGCC 1341
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTTCTGTGGATGGGCTACGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCCATCGAG 1401
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCCGAAGAGAGAGCTGACCGTGAACGACATCCGAAGCTGGTGGGCAAGCTGAAC 1461
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGCGCAGCCAGATCTACCCCGCATCAAGTGGCCGAGCTGCAAGCTCTCGCGGC 1521
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
QY 1522 GCCAAGGCCCTGACCGCATCGTCCCTGACCCGAGGAGGCGGAGCTGGAGCTGGCCGAG 1581
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1582 AACCGGAGATCTCGCGGAGCCCGTGCACCGGCTGTACTACGACCCCGAGCAAGAGACTG 1641
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480
QY 1642 GTGGCCGAGATCCAGAGCAGGGCCACGACAGTGGACCTACCAAGATCTACCGAGCC 1701
Db 481 IleAlaGluLeuGlnLysGlnGlyGlnGlyGlnTrpThrThrGlnIleTyrGlnGluPro 500
```

```
QY 1702 TTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGGCCACCGCCACACCAACGACGTG 1761
Db 501 PheLysIleLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCGAGGCGGTGCAGAAGATGCCATGCCATGAGAGCATCTGTGATCTGGGGCAAG 1821
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGlnSerIleValIleTrpGlyLys 540
QY 1822 ACCCCCAAGTTCGCCCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGACCGACTAC 1881
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1882 TGGCAGCGCCACTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1942 TGGTACCAGCTGGAGAAGGAGCCCATCATCGGCGCCGAGACCTTCTAGCTGGACGGGCC 2001
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCAGACCAAGATCGGCAAGCGCGGTACGTGACCCGACCGGGCGCGCAGAG 2061
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2062 ATCTGTAGCTGACCGAGACCCACCAACAGAGACCGAGCTGAGGCGCATCCAGCTGGCC 2121
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCAGGACGCGCAGCGAGGTGAACATCGTACCGACGACGACGAGTACGCGCTGGGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyLys 660
QY 2182 ATCCAGGCGCCAGCCGCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGACGAGCTG 2241
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAGGAGAGGTGTACTGTAGCTGGGTGGTCCCGCCCAAGAGGCGATCGGGCGAAC 2301
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
QY 2302 GAGCAGATCCAGACGAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTCCTCGACGGCATC 2361
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyLys 720
QY 2362 GAT 2364
Db 721 Asp 721
```

## RESULT 6

B44001

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)  
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly-  
C;Species: human immunodeficiency virus type 1, HIV-1

A;Note: host Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004

C;Accession: B44001

R;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.

J. Virol. 66, 6587-6600, 1992

A;Title: Complete nucleotide sequence, genome organization, and biological properties of

A;Reference number: A44001; MUID:93021387; PMID:1404605

A;Accession: B44001

A;Molecule type: DNA

A;Residues: 1-1003 <LIY>

A;Cross-references: UNIPROT:P35963; GB:M93258

C;Comment: This protein is synthesized as a gag-pol polyprotein.

C;Comment: Specific enzymatic cleavages may yield mature proteins including proteinase,

C;Genetics:

A;Gene: pol

C;Superfamily: pol polyprotein

C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo

F;57-155/Product: retropepsin #status predicted <RTP>

F;81/Active site: Asp (shared with dimeric partner) #status predicted

## Alignment Scores:

Pred. No.: 9,68e-153 Length: 1003  
Score: 3450.00 Matches: 647  
Percent Similarity: 94.87% Conservative: 37  
Best Local Similarity: 89.74% Mismatches: 31  
Query Match: 75.44% Indels: 6  
DB: 1 Gaps: 3

US-09-610-313B-31 (1-2463) x B44001 (1-1003)

QY 220 TTCTTCGCGAGACCTGGCTTCCCTCCCGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279  
Db 1 PhePheArgGluAepLeuAlaPheProGlnGlyLeuAlaArgLysPheSerSerGluGln 20  
QY 280 AACGGCCACACGCCCCACCAGCCGCGAGCTCGAGTG-----CGCGCGCAACCC 333  
Db 21 ThrArgAlaAsnSerProIleAArgGluArgGlnValTrpArgArgAspAsnAsnSer 40  
QY 334 CGCAGCGAGCCGCGCGCGAGCGCGAGCGCACCTCTG-----AACTTCCCGCCAGATCAC 387  
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
QY 388 CTGTGGCAGCGCCCTCTGTGTGAGATCAAGTGGGGCGGCAGATCAAGAGGCGCTGCTG 447  
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
QY 448 GACACGGCGCGCAGCACCGCTGCTGGAGGAGATGAGCTGCCCGCAAGTGAAGGCC 507  
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGGCGAGTACACAGACAGATCCTGATC 567  
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle 120  
QY 568 GAGATCTGGCGAAGAGCCATCGGCACCGCTGCTGATCGCGCCCGCCACCCCGTGAACATC 627  
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGGTGGCTGCACCTGACCTGAACTTCCCATCAGCCCATC 687  
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
QY 688 GAGACCGTGCCTGTGAAGTGAAGCCCGCATCGAGCGCGCCCAAGGTGAAGCAGTGGCCCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAAAGAGGGC 807  
Db 181 LeuThrGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AAGATCACCAAGATCGGCGCGGAGACCCCTTACACACCCCGCTGTTCGCCATCAAGAG 867  
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
QY 868 AAGGACACACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAAGCGCACCCAG 927  
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
QY 928 GACTTCTGGAGGTGACGTGGGCATCCCGCCACCGCGCGCTGGAAGAGAAAGAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
QY 988 GTACCGGTGCTGGACGTGGCGAGCGCTACTTACGGTGGCGCGCGAGGACTTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuHisGluAspPheArg 280  
QY 1048 AAGTACACCCCTTCCCATCTCCCGAGCATCAACAAAGAGACCCCGCGCATCCCGCTACCCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyThrArgTyrGln 300  
QY 1108 TACAAGTGTGCCCCAGGCTGGAGGGCAGGCCAGCATCTTCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

QY 1168 AAGATCTGTGAGCCCTTCGCGCGCCCAACCCGAGATCGTGATCTACCAG-----GCC 1221  
Db 321 ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTrpGlnTrpMetAsp 340  
QY 1222 CCCTGTAGCTGGGCGAGCACCTGGAGATCGGCGACGACCGCGCCCAAGATCGAGGAGCTG 1281  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1282 CGCAAGCACCTGTGGCTGGGCTTCACACCCCGCGCAAGAACAGACACAGAGAGAGCC 1341  
Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1342 CCCTTCTGTGATGGCTACGAGCTGACCGCACCGCAAGTGGACCTGCGAGCCCATCGAG 1401  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1402 CTGCCCGAGAAGAGAGCTGGACCGCTGAACACGACATCCAGAAGCTGGTGGCGAAGCTGAAC 1461  
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1462 TGGCCGAGCCAGATCTACCCCGGCATCAAGGTGCGGCACGCTGTGAAGCTGCTGCCGCGC 1521  
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
QY 1522 GCCAAGGCCCTGACCGACATCGTCCCTGACCGAGGAGCGGAGCTGGAGCTGGCGCGAG 1581  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuAlaGlu 460  
QY 1582 AACCGCGAGATCCTGCGGAGCCCGCTGACCGCGTGTACTACGACCCCGAGCAAGACCTG 1641  
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480  
QY 1642 GTGGCCGAGATCCAGAGCAGGCGCCACGACCGAGTGGACCTTACCAGATCTTACCAGAGCC 1701  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500  
QY 1702 TTCAAGAACTGAAGACCGCGCAAGTACGCAAGTGCACCGCCCGCCACACCAACGACGCTG 1761  
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520  
QY 1762 AAGCAGCTGACCGAGGCGCTGCAGAAAGATCGCCATGAGAGAGCATCTGTATCTGGGCGAAG 1821  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
QY 1822 ACCCCCAAGTTCGCTGCCATCCAGAGGAGACCTGGAGACCTGGTGGACCGACTAC 1881  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560  
QY 1882 TGGCAGGCCACCTGGATCCCGAGTGGGAGTTTCGTGAACACCCCGCTGTGTGAAGCTG 1941  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
QY 1942 TGGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGAGCGCGCC 2001  
Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600  
QY 2002 GCCAACCAGGAGACCAAGATCGGCAAGCGCGCTAGCTACCGACCGCGGCGCGGCGAAG 2061  
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnLysGlyArgGlnLys 620  
QY 2062 ATCTGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCCATCCAGCTGGCC 2121  
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640  
QY 2122 CTGCAGGACAGCGCGCAGAGGTGAACATCGTGAACCGACAGCAGCAGTACGCCCTGGGCATC 2181  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
QY 2182 ATCCAGGCCCGCGCAAGAGCGGAGCGAGCTGGTGAACAGAGATCATCGAGCAGCTG 2241  
Db 661 IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680



QY	1702	TTCAAGAACCTGAAGACCGCGCAAGTAGCCCAAGATCGCACCGCCACCAACGACGTG	1761
Db	501	PhelyAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal	520
QY	1762	AAGAGCTGACCGAGCGGTGCGAAGATGCCCATGGAGAGCATCGTGATCTGGGCAAG	1821
Db	521	LysGlnLeuThrGluAlaValGlnLysValSerThrGluSerIleValIleTrpGlyLys	540
QY	1822	ACCCCAAGTTCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTAC	1881
Db	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTrpMetGluTrp	560
QY	1882	TGGCAGCGCACCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGTGAAGCTG	1941
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu	580
QY	1942	TGTTACAGCTGAGAGAGGCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGGCC	2001
Db	581	TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	600
QY	2002	GCCAAACCGGAGACCAAGATCGGCAAGCGCGCTAGCTGACCGCGCGCGCAGAAG	2061
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys	620
QY	2062	ATCTGAGCTGACCGAGACCAACCAAGACCGAGCTGAGGCGCATCCAGCTGGCC	2121
Db	621	ValValSerIleAlaAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla	640
QY	2122	CTCAGAGACAGCGGCGAGGTGAACATCGTGACCGACGACGACGCTTACGCTGGGATC	2181
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
QY	2182	ATCCAGGCCAGCCGACAGAGCGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTG	2241
Db	661	IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu	680
QY	2242	ATCAAGAAGAGAGGTGTACCTGAGCTGGGTGCGCGCCGACCAAGGGCATCGCGCAAC	2301
Db	681	IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn	700
QY	2302	GACAGATCGAACGTGTGTGACGAAGGGCATCGGAAGGTGCTGTCTTCTGGACGGATC	2361
Db	701	GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAsnGlyIle	720
QY	2362	GAT 2364	
Db	721	Asp 721	
RESULT 8			
GNVWVL			
N;Contains: HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LV)			
C;Species: human immunodeficiency virus type 1, HIV-1			
C;Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 09-Jul-2004			
C;Accession: A03967			
R;Muesing, M.A.; Smith, D.H.; Cabradilla, C.D.; Benton, C.V.; Lasky, L.A.; Capon, D.J.			
Nature 313, 450-458, 1985			
A;Title: Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovi			
A;Reference number: A93355; PMID:85111157; PMID:2982104			
A;Molecule type: DNA			
A;Accession: A03967			
A;Cross-references: UNIPROT:P03368			
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re			
C;Genetics:			
A;Gene: pol			
C;Superfamily: pol polyprotein			
C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle			
F;66-164/Product: retropepsin #status predicted <RTP>			
F;90/Active site: Asp (shared with dimeric partner) #status experimental			
Alignment Scores:			
Pred. No.:	3.14e-152	Length:	1012

Score:	3439.00	Matches:	647
Percent Similarity:	93.67%	Conservative:	34
Best Local Similarity:	89.00%	Mismatches:	28
Query Match:	75.20%	Indels:	18
DB:	1	Gaps:	4
US-09-610-313B-31 (1-2463) x GNVWVL (1-1012)			
QY	238	GCCTTCCCCCAGGGCAAGCCCGCGAGTTC-----	267
Db	4	AlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGlnThrArgAlaAsnSerPro	23
QY	268	-----CCCAGCGAGCAGAACCCGCGCAACAGCCACCGCCAGCGGAGTGCAGGTG---	318
Db	24	ThrIleSerSerGluGlnThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrp	43
QY	319	----CGGGCGGACACCCCGCGAGCGGCGGGCGCGAGCGCCAGCGGCGCACCTTG-----	369
Db	44	GlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPhe	63
QY	370	AATTTTCCCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTGGCGGCCAG	429
Db	64	AsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGlyGln	83
QY	430	ATCAAGGAGGCGCTGTGTGACACCGCGCGCGACGACACCGCTGCTGGAGAGATGACCTG	489
Db	84	LeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeu	103
QY	490	CCCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGGCTTCATCAAGGTGCCCG	549
Db	104	ProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGln	123
QY	550	TACGACCATCTGTATCGAGATCTCGCGCAAGAGCGCATCGGCACCGCTGTGTATCGGC	609
Db	124	TyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIleGlyThrValLeuValGly	143
QY	610	CCACACCCCGTGAACATCATCGCGCGCAACATCTGTGACCGCTGGCTGCTGCACCTGAAC	669
Db	144	ProThrProValAsnIleIleGlyArgAsnLeuThrGlnIleGlyCysThrLeuAsn	163
QY	670	TTCCCATCAGCCCATCGAGACCGTCCCGGTGAGCTGAGCCCGGCGATGAGCGGCCCC	729
Db	164	PheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyPro	183
QY	730	AAGTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTGACCGCATCTCGCAG	789
Db	184	LysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThr	203
QY	790	GAGATGGAGAAGGAGGCAAGATCAACAAAGATCGGCGCCCGAGAACCCCTTACACACCCC	849
Db	204	GluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrPro	223
QY	850	GTGTTCCCATCAAGAAAGACACACCAAGTGGCGGCAAGCTGGTGAGCTTCCGCGAG	909
Db	224	ValPheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGlu	243
QY	910	CTGAACAAGCGACCCAGACACTTCTCGGAGGTGCAGCTGGGCATCCCCACACCCCGCGGC	969
Db	244	LeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGly	263
QY	970	CTGAAGAAGAAGAGCGTGCACCGTGTGGAGCGTGGCGGCGAGCCCTTCTACGCTGCC	1029
Db	264	LeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValPro	283
QY	1030	CTGGACGAGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACAGGACCC	1089
Db	284	LeuAspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnGluThr	303
QY	1090	CCCGGATCCGCTTACCAGTACAACTGCTGCCCGAGGTGGGAAGGCGGAGCCCGCAGATC	1149
Db	304	ProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIle	323
QY	1150	TTCCAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCGCAACCCCGAGATCGTG	1209

Db 324 PheGlnSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343  
Qy 1210 ATCTACCAG-----GCCCCCTGTGTACGTGGCGGACGACCTGGAGATCGGCCACACCGC 1263  
Db 344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363  
Qy 1264 GCCAAGATCGAGGAGCTGGCGAAGCAGCTGTGGCTGGGGCTTCACACCCCGACAG 1323  
Db 364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLys 383  
Qy 1324 AAGCACCAGAGAGCCCTCTCTGTGGATGGCTACGAGCTGCACCCCGACAAGTGG 1383  
Db 384 LysHisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrp 403  
Qy 1384 ACCGTGAGCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGACATCCAGAAG 1443  
Db 404 ThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLys 423  
Qy 1444 CTGTGGGCAAGCTGAATGGGCCAGGCAGATCTACCCCGGCATCAAGTGGCGCAGCTG 1503  
Db 424 LeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443  
Qy 1504 TGCAAGCTGTGGCGGCGCAAGCCCTGACCGACATCGTGGCCCTGACCGAGAGGCC 1563  
Db 444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463  
Qy 1564 GAGCTGAGCTGGCGGAGAACCGGACATCTCGCGGAGCCGCTGCACGGCGTGTACTAC 1623  
Db 464 GluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyr 483  
Qy 1624 GACCCAGCAGAGACTGTGGTGGCGAGATCCAGAACGAGGCGCACGACCAAGTGAACCTAC 1683  
Db 484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyr 503  
Qy 1684 CAGATCTACGAGGAGCCTTCAAGAACCTGAGACCGGACGATACGACCAAGTGGCACC 1743  
Db 504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523  
Qy 1744 GCCCACAACCAACGACGTGAAGCAGCTGACCGAGCGCGTGACAGAGATCGCCATCGAGC 1803  
Db 524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543  
Qy 1804 ATCGTGTATCTGGGCAAGACCCCAAGTTCGCTGCCCTCCATCCAGAGGAGACTGGAG 1863  
Db 544 IleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTrpGlu 563  
Qy 1864 ACCTGGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGGATTCGTGAACACC 1923  
Db 564 ThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThr 583  
Qy 1924 CCCCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGGCCCATCATCGGCGCGAGACC 1983  
Db 584 ProProLeuValLysLeuTyrTrpGlnLeuGlnLysGluProIleValGlyAlaGluThr 603  
Qy 1984 TTCTACGTGGACGGCGCGCCCAACCGCGAGACCAAGATCGCAAGCGCGGTACGTGACC 2043  
Db 604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623  
Qy 2044 GACCGGGCGCGAGAGATCGTGAAGCTGACCGAGACCCACCAAGACCGAGACCGAGCTG 2103  
Db 624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643  
Qy 2104 CAGGCCATCCAGCTGGCGCTGCAGACAGCGGACGAGGTGAACATCGTGACCGACAGC 2163  
Db 644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSer 663  
Qy 2164 CAGTACCGCTGGGCATCATCCAGCGCCAGCCGACAGAGCGAGAGCGAGCTGGTGAAC 2223  
Db 664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683  
Qy 2224 CAGATCATCGAGCAGCTGATCAGAAGAGAGAGTGTACTGAGCTGGTGGCGGCCAC 2283

Db 684 GlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProIleHis 703  
Qy 2284 AAGGGCATCGCGCGCAACGACGACATCACAAGCTGTGTAGCAAGGGCATCCGCAAGTGT 2343  
Db 704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723  
Qy 2344 CTGTTCCTTCGACGCGCATCGAT 2364  
Db 724 LeuPheLeuAspGlyIleAsp 730  
RESULT 9  
GNLJSI  
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz  
N;Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
C;Species: simian immunodeficiency virus SIVcpz  
A;Note: Host Pan troglodytes (chimpanzee)  
C;Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C;Accession: S09984  
R;Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A;Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A;Reference number: S09983; MUID:90259077; PMID:2188136  
A;Accession: S09984  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1027 <HUE>  
A;Cross-references: UNIPROT:P17283; EMBL:X52154  
C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
F;81-180/Product: retropepsin #status predicted <RTP>  
F;105/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
Pred. No.: 1,17e-141 Length: 1027  
Score: 3211.00 Matches: 595  
Percent Similarity: 88.92% Conservative: 63  
Best Local Similarity: 80.41% Mismatches: 77  
Query Match: 70.22% Indels: 5  
DB: 1 Gaps: 2  
US-09-610-313B-31 (1-2463) x GNLJSI (1-1027)

Qy 159 CTGCTGGAAGTGGCGCAAGCGAGCGCCACCATGAGGACTGCACCGAGCGCCAGGCCAA 218  
Db 7 LeuLeuAlaValTrpAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26  
Qy 219 CTTTCTTCGCGAGGACTGCGCTTCCCGGAGGCAAGCGCGCGAGTTCCCGACGAGCA 278  
Db 27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGluG 46  
Qy 279 GAACCGCGCCACACGCCCCACCGAGCGCGAGCTGCAGTGGCGCGC-----GACAACCC 332  
Db 46 nAsnArgThrAsnGlyProThrAspArgGluLeuTrpValProGlyGlyArgGluGluPr 66  
Qy 333 CCGCAGGAGCGCGCGCGCGCGAGCGCGCGCCCTCAACTTCCCGCAGATCACCCCTGTG 392  
Db 66 oGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86  
Qy 393 GCAGCGCCCCCTGGTGAAGCATCAAGTGGCGCGCGAGATCAAGGAGGCCCTCTGTGGACAC 452  
Db 86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106  
Qy 453 CGGCGCGCGACACACCGTGTGGAGGAGATGAGCTCTCCCGGCAAGTGGAGGCCCAAGAT 512  
Db 106 rGlyAlaAspAspThrValIleGluArgIleGlnLeuGlnGlyLeuTrpLysProLysMe 126  
Qy 513 GATCGGCGGATCGCGCGCTTCATCAAGGTGGCGCGAGTACGACAGATCCTGTATCGAGAT 572  
Db 126 ttleGlyGlyIleGlyPheIleLysValLysGlnPheAspAsnValHisIleGluI 146





```
QY 481 ATGAGCTGTCGCGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTCATCAAG 540
DB 1 MetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 20
QY 541 GTGCGCCAGTACGACAGATCCTGATCAGATCTGCGCGCAAGAGGCATCGGCACCGTG 600
DB 21 ValysGlnTrpAspGlnIleAlaIleGluLeuCysGlyHisLysAlaIleGlyThrVal 40
QY 601 CTGATCGGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
DB 41 LeuValGlyProThrProValAsnIleGlyArgAsnLeuLeuLeuThrGlnLeuGlyCys 60
QY 661 ACCCTGAACCTCCCATCGACCCCATCGACACCGTGCCTGCGTGAAGCTGAAGCCCGGATG 720
DB 61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80
QY 721 GACGGCCCCAAGTGAAGCAGTGCCTCCACGAGGAGAGATCAAGGCGCTGACCGCC 780
DB 81 AspGlyProLysValLysGlnTrpProLeuThrGlnLysIleLysAlaLeuIleGlu 100
QY 781 ATCTGCGAGGAGATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTAC 840
DB 101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120
QY 841 AACACCCCGCTGTTCGCCATCAAGAGAAGAGACAGCACCAAGTGCAGCAAGCTGGTGGAC 900
DB 121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140
QY 901 TTCGCGGAGCTGAACAGCCACCCAGGACTCTCGGAGGTGCAGCTGGGCGATCCCCAC 960
DB 141 PheArgGlnLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160
QY 961 CCGCGCGCTGGAAGAGAAGAGAGCGTCAACGCTGCTGGACGTGGCGCAGCGCTACTTC 1020
DB 161 ProGlyGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180
QY 1021 AGCGTGCCCTCGACGAGGACTTCGCAAGTACACCGCTTCACCATCCCGAGATCAAC 1080
DB 181 SerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsn 200
QY 1081 AACGAGACCCCGGATCCCTACCACTACAGTACAAGTGTCTGCCAGGCGTGGAGGGCAGC 1140
DB 201 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 220
QY 1141 CCAGAGATCTTCAGACGACATGACCAAGATCCTGGAGCCCTTCGCGCGCCGCAACCCC 1200
DB 221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240
QY 1201 GAGATCGTGATCTACCAAG-----GCCCGCTGTACGTGGCGACGACCTGGAGATCGGC 1254
DB 241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260
QY 1255 CAGCAGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGGCTGGGCTTCACCAAC 1314
DB 261 GlnHisArgThrLysIleGluLeuArgArgHisLeuLeuArgTrpGlyPheThrThr 280
QY 1315 CCGAGCAAGAGCAGCAGGAGGAGCCCTCTCTGTGGATGGCTACGAGCTGCACCCC 1374
DB 281 ProAspLysLysHisGlnLysGluProPheLeuTrpIleGlyTyrGluLeuHisPro 300
QY 1375 GACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGTGAACGAC 1434
DB 301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320
QY 1435 ATCCAGAGCTGGTGGCAAGCTGAAGTGGCGCAGCCAGATCTACCCCGGCATCAAGGTG 1494
DB 321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340
QY 1495 CGCCAGCTGTGAAGCTGCTGCGCGCGCAAGCCCTGACCGACATCGTGGCCCTCGAC 1554
DB 341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360
QY 1555 GAGGAGGCGCGAGCTGGGCGCGAGAACCGGAGATCCTGCGCGAGCCCGTGCACGCG 1614
```

```
DB 361 LysGluAlaGlnLeuGluLeuAlaGluAsnArgGluIleLeuLysThrProValHisGly 380
QY 1615 GTGTACTTACGACCCCAAGAGACCTGCTGCCGAGATCCAGAAGCAGGGGCCACGACCAG 1674
DB 381 ValTyrTyrAspProSerLysAspLeuValAlaGluIleGlnLysGlnGlyLeuGlyGln 400
QY 1675 TGGACCTTACAGATCTTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1734
DB 401 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420
QY 1735 ATGGGACCCCGCCACACCAAGCAGCTGAACAGCTGACCGGCGCTGCGAGAGATCGCC 1794
DB 421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440
QY 1795 ATGAGAGATCTGTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAG 1854
DB 441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460
QY 1855 ACCTGGAGACCTGGTGACCGACTACTGCGAGCCACCTGGATCCCCGAGTGGAGTTC 1914
DB 461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480
QY 1915 GTGAACACCCCGCTGGTGAAGCTGTGTACAGCTGGAGAGGAGGCCCATCATCGGC 1974
DB 481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500
QY 1975 GCCGAGACCTTCTACGTGGACGCGCCCAACCGCAGACCAAGATCGGCAAGCCCGGC 2034
DB 501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520
QY 2035 TACGTGACCGACCGGCGCGCAGAGATCGTGAAGCTGACCGCTGACCGACCAACCAAG 2094
DB 521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys 540
QY 2095 ACCGAGCTGACGCGCATCCAGCTGGCCCTCGCAGACAGCGCAGCGAGTGAACATCGTG 2154
DB 541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560
QY 2155 ACCGACAGCCAGTACGCGCTGGGCGCATCATCCAGCCCGCAGCCCGCAAGAGCGAGCGAG 2214
DB 561 ThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnProAspLysSerGluSerGlu 580
QY 2215 CTGGTGAACAGATCATCGACGAGCTGATCAAGAGAGAGAGTGTACTCTGAGCTGGGTG 2274
DB 581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600
QY 2275 CCGCGCCCAAGGCGCATCGCGCGCAGCAGCAGATCGACAGCTGGTGGAGCAAGGGCATC 2334
DB 601 ProAlaHisLysGlyIleGlyGlyAsnGluValAspLysLeuValSerAlaGlyIle 620
QY 2335 CGCAAGGTGCTGTCTCTGGACGCGCATCGAT 2364
DB 621 ArgLysValLeuLeuAspGlyIleAsp 630

RESULT 11
T01668
pol polyprotein - human immunodeficiency virus type 1
C;Species: human immunodeficiency virus type 1, HIV-1
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T01668
R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.
Cell 46, 63-74, 1986
A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol
A;Reference number: Z14389; MUID:86245056; PMID:2424612
A;Accession: T01668
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-902 <Al>
A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230
C;Superfamily: pol polyprotein

Alignment Scores:
```

Pred. No.: 3,05e-136 Length: 902  
Score: 3094.00 Matches: 574  
Percent Similarity: 97.10% Conservative: 28  
Best Local Similarity: 92.58% Mismatches: 16  
Query Match: 67.66% Indels: 2  
DB: 2 Gaps: 1

US-09-610-313B-31 (1-2463) x T01668 (1-902)

QY 511 ATGATCGGGGATCGGCGCTTCATCAAGTGGCCAGTACGACAGATCCTGATCGAG 570  
Db 1 MetIleGlyGlyIleGlyPheIleLysValArgInTrpAspGlnIleLeuIleGlu 20  
QY 571 ATCTGGGCAAGAGCCATCGGCACCGTGTGATCGGCCCCACCCCGTGAACATCATC 630  
Db 21 IleCysGlyLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 40  
QY 631 GCGCGCAACATGTGACCCAGCTGGGTGCGACCTGAATCTCCCATCAGCCCCATCGAG 690  
Db 41 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 60  
QY 691 ACCGTGCCGTGAAGCCGCGCATGGAGCGGCCCAAGGTGAAGCAGTGGCCCTCG 750  
Db 61 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 80  
QY 751 ACCGAGGAGAATCAAGGCCCTGACCGGCATCTGCGAGGATGAGAGAGGGCAAG 810  
Db 81 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluLys 100  
QY 811 ATCACCAAGATCGGCCGAGAACCCCTACAACACCCCGTGTTCGCCATCAAGAGAAG 870  
Db 101 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 120  
QY 871 GACAGACCAAGTGGCGGCACTGGTGGACTTCCGCGAGCTGACAGCGCACCCAGGAC 930  
Db 121 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 140  
QY 931 TTCTGGAGGTGAGTGGCGATCCCCACCCCGCGCTGAAGAGAAGAGAGCGGTG 990  
Db 141 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 160  
QY 991 ACCGTCTGGAGTGGGCAAGCCTACTTCAGCGTCCCTGAGCAGACTTCCGCAAG 1050  
Db 161 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 180  
QY 1051 TACACCGCTTACCATCCCGCATCAACAGAGACCCCGCATCCGCTACCAAGTAC 1110  
Db 181 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 200  
QY 1111 AACGTGCTGCGCGCTGGAGGGCAGCCCGCATCTTCCAGAGCAGCATGACCAAG 1170  
Db 201 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 220  
QY 1171 ATCTCGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCTACAG-----GCCCCC 1224  
Db 221 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 240  
QY 1225 CTGTAGCTGGGCGACCTGGAGATCGCCAGCACCGCGCCCAAGATCGAGGAGCTGGC 1284  
Db 241 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 260  
QY 1285 AAGCACTGTGCGTGGGCTTCCACACCCCGCAAGAAGCACCAGAGAGAGCCCGCC 1344  
Db 261 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 280  
QY 1345 TTCTCTGGATGGCTTACGAGTGCACCCCGCAAGTGAACCGTGCAGCCCATCGAGCTG 1404  
Db 281 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 300  
QY 1405 CCGGAGAAGAGCTGGACCGTGAAGCATCCAGACATCCAGAGCTGGTGGCAGCTGAAGT 1464  
Db 301 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 320

QY 1465 GCCAGCCAGATCTACCCCGCATCAAGGTGCGCCAGCTGTGCAAGCTGTCGCGCGCGCC 1524  
Db 321 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 340  
QY 1525 AAGGCCCTGACGACATCGTCCCTGACCGAGGAGGCGAGCTGGAGCTGGCCGAGAAC 1584  
Db 341 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 360  
QY 1585 CCGGAGATCCTCGCGAGCCCGTGCACGCGCTGTACTACGACCCCAAGACCTGGTG 1644  
Db 361 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 380  
QY 1645 GCGGAGATCCAGAAGCGGCCACGACCTGAGACCTACAGATCTACAGAGACCCCTTC 1704  
Db 381 AlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluGlnTyr 400  
QY 1705 AAGAACCTGAACGCGCAAGTACCGAAGATGCGCAGATGCGCAGCCGACACCAACGACGTGAAG 1764  
Db 401 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 420  
QY 1765 CAGCTCACCGAGCGCGTGCAGAAGATCGCCATGAGAGATCGTGTATCTCGGGCGAAGACC 1824  
Db 421 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 440  
QY 1825 CCCAAGTTCGCTGCCATCCAGAGGAGACTGGAGACCTGGTGGACCGACTACTGG 1884  
Db 441 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 460  
QY 1885 CAGGCCACCTGATCCCGAGTGGGAGTTCGTGACACACCCCGCTGGTGAAGCTGG 1944  
Db 461 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 480  
QY 1945 TACCAGTCGAGAAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGACCGCGCGCC 2004  
Db 481 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 500  
QY 2005 AACCCGAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGCGCGCGAGATC 2064  
Db 501 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 520  
QY 2065 GTGAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCGAGCCATCCAGCTGGCCCTG 2124  
Db 521 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 540  
QY 2125 CAGGACAGCGCGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTCGGCATCATC 2184  
Db 541 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 560  
QY 2185 CAGGCCACCGCCGACAAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGAGCTGATC 2244  
Db 561 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 580  
QY 2245 AAGAAGGAGAGGTGTACCTGAGCTGGGTGCGCCGCCCAAGGGCATCGCGCAACGAG 2304  
Db 581 GlnLysAspLysValTyrLeuSerTrpValProAlaHisGlyGlyIleGlyGlyAsnGlu 600  
QY 2305 CAGATCGACAAGCTGTGTAGCAAGGATCCCGAAGGTGCTTCTCGACCGCATCGAT 2364  
Db 601 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 620

RESULT 12  
B47175

reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1  
C:Species: human immunodeficiency virus type 1, HIV-1  
C>Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: B47175  
R:Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.  
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993  
A:Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the  
A:Reference number: A47175; MUID:93126353; PMID:7678340  
A:Accession: B47175  
A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: DNA

A;Residues: 1-559 <MOH>  
A;Cross-references: UNIPROT:Q9PXX1  
A;Note: sequence extracted from NCBI backbone (NCBIP:122099)  
C;Superfamily: pol polyprotein

## Alignment Scores:

Pred. No.:	2,69e-121	Length:	559
Score:	2771.00	Matches:	514
Percent Similarity:	96.06%	Conservative:	23
Best Local Similarity:	91.95%	Mismatches:	20
Query Match:	60.59%	Indels:	2
DB:	2	Gaps:	1

US-09-610-313B-31 (1-2463) x B47175 (1-559)

```
QY 673 CCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGACGCGCCCAAG 732
Db 1 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 20
QY 733 GTGAAGCAGTGGCCCTGACCGAGGAGAGATCAAGAGCCCTGACCGCCATCTCGGAGGAG 792
Db 21 ValLysGlnTrpProLeuThrGluGluLysAlaLeuValGluLysCysThrGlu 40
QY 793 ATGAGAGAGGAGGAGATCACCAAGATCGCCCGAGAACCCCTACACACCCCGCTG 852
Db 41 MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal 60
QY 853 TTCGCCATCAAGAAGAGGAGCAGCACCAAGTGGCGCAAGCTGGTGAATTCCTCGCGAGCTG 912
Db 61 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 80
QY 913 AACAAAGCGCACCGAGACTTCTGGAGGTGCAGCTGGGCATCCCGACCCCGCGGCTG 972
Db 81 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 100
QY 973 AACAAAGAGAGCGTGCAGCTGGAGCTGGAGCTGGGCGACGCTTACGCTGCCCTG 1032
Db 101 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 120
QY 1033 GACGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGAGCATCAACACGAGACCC 1092
Db 121 AspGluAspPheArgLysThrAlaPheThrIleProSerIleAsnAsnGluThrPro 140
QY 1093 GGATCCGCTACCAAGTACACGCTGCTGCCCGAGGCTGGAAGGCGACGCCAGCATCTTC 1152
Db 141 GlyLeuArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 160
QY 1153 CAGAGCAGATGACCAAGATCTCTGGAGCCCTTCGGCGCCCGCAACCCCGAGATCGTGATC 1212
Db 161 GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIle 180
QY 1213 TACCAG-----GCCCGCTGTAGTGGGCGAGGACCTGGAGATCGGCGCAGCACCGCGCC 1266
Db 181 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLysGlyGlnHisArgThr 200
QY 1267 AAGATCGAGGAGCTGCGCAAGCAGCTGCTCGCTGGGGCTTACCAACCCCGCAAGAAG 1326
Db 201 LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys 220
QY 1327 CACGAGAGGAGCCCTTCCTGTTGATGGGTGATGAGTGCACCGCCCGCAAGTGGAC 1386
Db 221 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 240
QY 1387 GTCCAGCCCATCGAGCTGCCGAGAGGAGGAGTGGACCGTGAACGATCCAGAGAGCTG 1446
Db 241 ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu 260
QY 1447 GTGGGCAAGTGAATCGGGCAGCAGCATATACCCCGCATCAAGTGGCGCGAGCTGTGC 1506
Db 261 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuCys 280
QY 1507 AAGCTGCTGGCGCGCCAAAGGCCCTGACCGATCGTCCCTGACCGGAGGAGCGGAG 1566
Db 1566
```

```
Db 281 LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGlu 300
QY 1567 CTGAGCTGGCCGAGAACCCGAGATCTCTCGCGAGCCCGTGCACGCGTGTACTACGAC 1626
Db 301 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 320
QY 1627 CCCAGCAAGACCTGTGGCGGAGATCCAGAACGAGGCGCCACGACGAGTGGACCTAC 1686
Db 321 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGln 340
QY 1687 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCAAGATGCGCACCGCC 1746
Db 341 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla 360
QY 1747 CACACCAACGCTGAAGCAGCTGACGAGCGCGTGCAGAAATCGCCATCGGAGAGCATC 1806
Db 361 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle 380
QY 1807 GTGATCTGGGCAAGACCCCAAGTTCGCGCTCCCATCCAGAGGAGACCTGGAGAC 1866
Db 381 ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr 400
QY 1867 TGGTGACCCGACTACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCC 1926
Db 401 TrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 420
QY 1927 CCCTCGTGAAGCTGTGTACCGAGCTGGAGAGAGCCCATCATCGCGCGGAGACCTTC 1986
Db 421 ProLeuValLysLeuTrpTyrGlnLeuLysGluProIleValGlyAlaGluThrPhe 440
QY 1987 TAGCTGAGCGGCGCCGCAACCCGAGACCAAGATCGGCAAGGCGCGCTACGTGACCGAC 2046
Db 441 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 460
QY 2047 CGGCGCGCGCAGAGAGTCTGTGACCTGACCGAGACCAACAGAGAGACCGAGCTGCAG 2106
Db 461 LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln 480
QY 2107 GCATCCAGCTGGCCCTTCGAGGAGCAGCGCAGGAGTGAACATCGTACCGACGAGCCAG 2166
Db 481 AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln 500
QY 2167 TAGCCCTGGGCATCATCGAGCCCGCAGCCGACGAGGAGGAGGAGGAGTGGTGAACCGAG 2226
Db 501 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln 520
QY 2227 ATCATCGAGCAGCTGATCAAGAGGAGAGTGTACCTGAGCTGGGTGGCGCCCAACAG 2286
Db 521 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys 540
QY 2287 GGATCGGCGGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTG 2343
Db 541 GlyIleGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal 559
```

## RESULT 13

A47175  
Reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1  
C;Species: human immunodeficiency virus type 1, HIV-1  
C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C;Accession: A47175  
R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.  
Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993  
A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the  
A;Reference number: A47175; PMID:93126353; PMID:7678340  
A;Accession: A47175  
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A;Molecule type: nucleic acid  
A;Residues: 1-559 <MOH>  
A;Cross-references: UNIPROT:Q9PXX2  
A;Note: sequence extracted from NCBI backbone (NCBIP:122100)  
C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	2.28e-120	Length:	559
Score:	2751.00	Matches:	510
Percent Similarity:	95.89%	Conservative:	26
Best Local Similarity:	91.23%	Mismatches:	21
Query Match:	60.16%	Indels:	2
DB:	2	Gaps:	1
US-09-610-313B-31 (1-2463) x A47175 (1-559)			
QY	673	CCCATCAGCCCATCAGAGCCGTCGCGTGAAGCTGAAGCCCGGCATGAGCGGCCCAAG	732
Db	1	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	20
QY	733	GTGAAGCAGTGGCCCTCGACCGAGGAGAAGATCAAGGCCCTGACCCCATCTCCGAGGAG	792
Db	21	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu	40
QY	793	ATGGAGAAGAGGGCAAGATCACCAAGATCGGCCCGCCGAGAACCCCTACAAACCCCGCTG	852
Db	41	MetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal	60
QY	853	TTCCGCATCAAGAAGAGGACACCAAGTGGCGCAAGCTGGTGACTTCGCGAGCTG	912
Db	61	PheAlaIleLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu	80
QY	913	AACAAGCGCACCCAGGACTTCTGGGAGTGCAGCTGGGCATCCCCCAGCCCGGCCCTG	972
Db	81	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	100
QY	973	AGAAGAAGAGAGCTGACCGTGTGGAGTGGAGCGGCGAGCGCTTACCTGAGCTGCCCTG	1032
Db	101	LysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	120
QY	1033	GACGAGGACTTCGCGAAGTACACCGCTTCACCATCCCCAGCATCAACACGAGACCCCC	1092
Db	121	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	140
QY	1093	GGCATCCGCTACCAAGTACAGCTGTGCCCCGAGGGCTGGAAGGGCAGCCCAACATCTTC	1152
Db	141	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	160
QY	1153	CAGAGCAGATGACCAAGATCTCTGGAGCCCTTCGCGCCCGGCACACCCGAGATCGTGATC	1212
Db	161	GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle	180
QY	1213	TACCAG-----GCCCCCTGTACGTGGCAGCGACCTCGAGATCGCCAGCACCGCGCC	1266
Db	181	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr	200
QY	1267	AAGATCGAGGAGTGGCGCAAGCACCTGTGCTCGCTGGGGCTTCAACACCCCGCACAGAAG	1326
Db	201	LysIleGluLeuLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	220
QY	1327	CACCAAGAAGAGCCCTCTCTCTGTGGATGGGTACGAGCTGACCCCGCACAAAGTGACCC	1386
Db	221	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	240
QY	1387	GTGCAGCCCATCGAGTGGCCGAGAGGAGAGCTGGACCGTGAACCATCCAGACAGCTG	1446
Db	241	ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu	260
QY	1447	GTGGGCAAGCTGAACCTGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCGAGCTGTGC	1506
Db	261	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValLysGlnLeuCys	280
QY	1507	AAGCTCTCGCGGCCCAAGGCCCTTGACCGACATCGTCCCTGACCGAGAGGCCGAG	1566
Db	281	LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluAlaGlu	300
QY	1567	CTGGAGCTGGCAGAACCGCAGATCTTGGCGAGCCGTGCACGGCTGTACTACGAC	1626
Db	301	LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp	320

QY	1627	CCCAGCAACGAGCCTGTGTGGCCGAGATCCAGAGCAGGGGCCACAGCACGATGGACCTACAG	1686
Db	321	ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGln	340
QY	1687	ATCTACCAAGAGCCCTTCAAGAACCTGAGACCGCGCAGTACGCCAAGATGGCACCAGCC	1746
Db	341	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla	360
QY	1747	CACACCAACGAGCTGAAGCAGCTGACCGAGGCGGTGCAGAAATGCCATCGAGAGCATC	1806
Db	361	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle	380
QY	1807	GTGATCTGGGCAAGACCCCAAGTTCCGCCTGCCCATCCAGAGAGACCTGGGAGACC	1866
Db	381	ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla	400
QY	1867	TGGTGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCC	1926
Db	401	TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	420
QY	1927	CCCTTGGTGAAGCTGTGTACCAGCTGGAGAAAGGAGCCCATCATCGGCCCGAGACTTC	1986
Db	421	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe	440
QY	1987	TACGTGGACGGCGCCCAACCCGAGACCAAGATCGGCAAGCCCGCTACGTGACCCAG	2046
Db	441	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	460
QY	2047	CGGGCGCGCAGAAATCGTGAGCTGACCGAGACCCAGAGACCAACACGAGACCCGCTG	2106
Db	461	LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln	480
QY	2107	GCCATCCAGCTGGCCCTGCAGGACAGCGCGAGTGAACATCGTGACCCAGACGCCAG	2166
Db	481	AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln	500
QY	2167	TACGCCCTGGGCATCATCCAGCCCGCCAGCAAGAGCAGAGCGAGCTGTGAACCCAG	2226
Db	501	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln	520
QY	2227	ATCATCGACGACTCATCAGAGGAGAGGTGTACCTGAGCTGGTGGTCCCCCCCCACAG	2286
Db	521	IleIleGluGluLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys	540
QY	2287	GGCATCGCGCGCAACGAGCAGATCGACAGCTGTGTGAGCAAGCGCATCGCAAGGTG	2343
Db	541	GlyIleGlyGlyAsnGluValValAspLysLeuValSerAlaGlyIleArgLysVal	559
RESULT 14			
S46347			
pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)			
C:Species: simian immunodeficiency virus SIVagm			
A:Variety: isolate SAB-1			
C:Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999			
R:Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.; et			
EMBO J. 13, 2935-2947, 1994			
A:Title: Mosaic genome structure of simian immunodeficiency virus from West African green			
A:Reference number: S46335; MUID:94298785; PMID:8026477			
A:Accession: S46347			
A:Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-1039 <JTN>			
A:Cross-references: EMBL:U04005; NID:9466229; PIDN:AAA21505.1; PID:9466231			
A:Experimental source: isolate SAB-1; sabaues monkey			
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993			
A:Note: this reading frame extends between two stop codons and does not begin with a start			
C:Genetics:			
A:Gene: pol			
C:Superfamily: pol polyprotein			
Alignment Scores:			
Pred. No.:	9.73e-102	Length:	1039

Score: 2350.00 Matches: 440  
Percent Similarity: 74.07% Conservative: 120  
Best Local Similarity: 58.20% Mismatches: 152  
Query Match: 51.39% Indels: 44  
DB: 2 Gaps: 7

US-09-610-313B-31 (1-2463) x S46347 (1-1039)

QY 220 TTCTTCGCGAGACCTGGCTTCCCCAGGCGC-----AAGGCCCGCGATTCCCGCAGC 273  
Db 1 PhePheArg-----ValTrpProLeuGlyGlnArgGluThrGlnGluPheProSer 17  
QY 274 GAGCAGAACCGCGCCAC-----AGCCCCACAGCGCGAGTGCAGGTGCGCGC----- 324  
Db 18 AspLeuHisGlnThrAsnSerProAsnGlyThrGlyLeuGlnAlaGlyGlyLys 37  
QY 325 -----GACACCCCCCGCAGGAGGCC----- 345  
Db 38 LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgSerAsnSerPro 57  
QY 346 -----GGCGCCGAGCGC----- 357  
Db 58 ValLysAlaValCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla 77  
QY 358 -----CAGGCGACCTGAACCTTCCCCAGATCACCTGTGCGCAGCGCCC 402  
Db 78 ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgArgPro 97  
QY 403 CTGGTGAGCATCAAGTGGCGCGCCAGATCAAGGAGCCCTGTGGACACCGCGCGCAG 462  
Db 98 MetCysThrValTyrIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp 117  
QY 463 GACACCGTGTGGAGGAGATGAGCTGCCCGCGCAAGTGGAGGCCCAAGATGATCGCGCGC 522  
Db 118 AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly 137  
QY 523 ATCGCGCGCTTCAAGTGGCGCGCAGTAGACACAGATCCTGTGATCGAGATCTCGGCGAAG 582  
Db 138 IleGlyGlyCysIleAsnValLysAlaTyrHisAsnGlnGluValLysIleGluAspLys 157  
QY 583 AAGGCCATCGGCGCGTGTGATCGCGCCCGCCCGTGAACATCATCGCGCGCCACATG 642  
Db 158 ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal 177  
QY 643 CTGACCCAGCTGGCGTGCACCTGAACTTCCCATCAGCCCCATCGAGACCGTGCCTCGT 702  
Db 178 LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal 197  
QY 703 AAGCTGAAGCCCGGCATGAGCGGCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAG 762  
Db 198 HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys 217  
QY 763 ATCAAGCCCTGACCGCATCTGGAGGAGATGGAGNAGGAGGCGCAAGATCACCAAGATC 822  
Db 218 IleGluAlaLeuLysAlaIleCysGluAspLeuGluLysGlnGlyHisLeuGluArgIle 237  
QY 823 GGCCCGGAGAACCCCTACAACACCCCGTGTGCCATCAAGAGAGAGGACAGCACCAAG 882  
Db 238 GlyProGluAsnProTyrAsnThrProValPheAlaIleArgLysLysAspLysThrGln 257  
QY 883 TGGCGCAAGCTGTGGACTTCCCGAGCTCAACAGCGCACCCAGGACTTCTGGGAGGTG 942  
Db 258 TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal 277  
QY 943 CAGCTGGGCATCCCGCCCGCGCTGAAGNAGNAGAGCGTGACCGTGGTGGAC 1002  
Db 278 GlnLeuGlyIleProHisProAlaGlyLeuGlnGlnArgGlnGlnIleThrValLeuAsp 297  
QY 1003 GTGGCGGACCGCTTACTTCCAGTGGCGCTCGAGAGGACTTCCGCAAGTACACCGCTTC 1062  
Db 298 IleGlyAspAlaTyrPheSerCysProLeuAspProAspPheGlnLysTyrThrAlaPhe 317  
QY 1063 ACCATCCCCAGCATCAACACGAGACCCCGCGCATCCGCTTACCAGTACAAACGTGCTGCC 1122

Db 318 ThrIleProSerValAsnAsnArgGluProGlyIleArgTyrGlnTyrLysValLeuPro 337  
QY 1123 CAGGGCTGGAAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCCTCGAGGCC 1182  
Db 338 GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu 357  
QY 1183 TTCGGCGCCCGCAACCCCGAGATCGTATCTACAG-----GCCCCCTGTAGTGGCGC 1236  
Db 358 PheArgGlnLysAsnProAspValAspIleTyrGlnTyrMetAspAspMetLeuIleAla 377  
QY 1237 AGGACCTGGAGATCGGCGAGCAGCGCCCAAGATCGAGAGCTGCCCAAGACCTGCTG 1296  
Db 378 SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTyrLeuGlu 397  
QY 1297 CGCTGGCGCTTCCACCCCGCAGCAAGACCCAGAGAGCGCCCTTCTCTGTGATG 1356  
Db 398 ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProProTyrLeuTrpMet 417  
QY 1357 GGCTACGAGTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGGAG 1416  
Db 418 GlyTyrGluLeuTyrProLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu 437  
QY 1417 AGCTGGACCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCAGCAGTC 1476  
Db 438 GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 457  
QY 1477 TACCCCGGATCAAGTGGCGCAGCTGTGCAAGCTGTGGCGCGCGCAAGGCCCTGACC 1536  
Db 458 TyrThrGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyAlaArgProLeuThr 477  
QY 1537 GACATCGTGCCTGACCCGAGGCGCGAGCTGCAGCTGCCCGAGAACCGCGAGATCTG 1596  
Db 478 GluIleValGlnTrpThrGluGluAlaGluLeuGluLeuGluAsnArgGlnIleLeu 497  
QY 1597 CGCGAGCCGTGACCGCGTGTACTAGCACCCAGCAAGACCTGGTGGCGAGATCCAG 1656  
Db 498 ArgGlnLysGlnGlnGlyGlnTyrAspProAlaLeuProLeuArgAlaLysValLeu 517  
QY 1657 AAGCAGGCCCGCAGCAGTGGACCTACACAGATCTACCGAGCGCTTCAAGAACCTGAAG 1716  
Db 518 LysLeuGlyAspGlyGlnTrpGlyTyrGlnIleTyrGlnProGluAsnLysIleLeuLys 537  
QY 1717 ACCGCGAAGTACCCCAAGATGCGCAGCGCCCAACCAACGAGTGAAGCTGACCGAG 1776  
Db 538 ValGlyLysTyrAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly 557  
QY 1777 GCGTGCAGAGATCGCCATGGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCGCG 1836  
Db 558 LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu 577  
QY 1837 CTGCCCATCCAGAAAGGAGACCTGGGAGACCTGTGGACCGACTACTGGCAGGCCACCTGG 1896  
Db 578 LeuProValGluArgGluLeuTrpGluTrpSerAspTyrTrpGlnValThrTrp 597  
QY 1897 ATCCCGAGTGGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGCTGCTGAGTGGAG 1956  
Db 598 IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTyrLysLeuVal 617  
QY 1957 AAGGAGCCCATCATCGCGCGCGAGACCTTCTAGTGGACCGCGCGCCCAACCGCGAGACC 2016  
Db 618 LysAspProIleProGlyGluAlaValTyrValAspGlyAlaAlaAsnArgAsnSer 637  
QY 2017 AAGCTGGCAAGCGCGCTACGTGACCGCGCGCGCGCAGAGATCGTGAGCTTCGAGCCT 2076  
Db 638 LysGluGlyLysAlaGlyTyrLeuThrAspArgGlyAspGlnLysValAlaLeuGlu 657  
QY 2077 GAGACCAACCAAGCAGAGCAGCTGCAGCGCCATCCAGCTGGCGCTGCGAGCAGCGCGC 2136  
Db 658 AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly 677  
QY 2137 AGCAGGTGAACATCGTGACCGCAGCAGCTAGCCCTGGCGCATCATCCAGGCCCGCACC 2196

Db 678 SerLysValAsnIleIleThrAspSerGlnTyrAlaMetGlyIleAlaGlyLeuPro 697  
QY 2197 GACAGAGCGAGCGAGCTGTGTGAACCAAGATCATCGACAGCTGATCAAGAGGAGAG 2256  
Db 698 ThrGluSerAsnAsnIleValGlnGlnIleIleGluLeuIleIleLysGluAla 717  
QY 2257 GTGTACTCTGAGTGGTGGCCGCCCAAGGGGATCGCGGGAACAGAGCAGATCGACAAG 2316  
Db 718 ValTyrIleAlaTrpValProAlaHisLysGlyValGlyAsnGluLeuIleAspLys 737  
QY 2317 CTGGTGAGCAAGGCGATCCGCAAGTGTCTTCCTGGACGGCATCGAT 2364  
Db 738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753

RESULT 15  
S53092  
pol polyprotein - human immunodeficiency virus type 2  
C:Species: human immunodeficiency virus type 2, HIV-2  
C:Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C:Accession: S53092  
R:Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.  
submitted to the EMBL data library, March 1995  
A:Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu  
A:Reference number: S53091  
A:Accession: S53092  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1055 <BEC>  
A:Cross-references: UNIPROT:O73194; EMBL:Z48731  
C:Superfamily: pol polyprotein  
C:Keywords: polyprotein

Alignment Scores:  
Pred. No.: 3.36e-95 Length: 1055  
Score: 2209.00 Matches: 417  
Percent Similarity: 70.60% Conservative: 128  
Best Local Similarity: 54.02% Mismatches: 186  
Query Match: 48.31% Indels: 41  
DB: 2 Gaps: 10

US-09-610-313b-31 (1-2463) x S53092 (1-1055)

QY 156 GGGCTGCTGGAAGTGGCGGAAGGAGGGCCACCAGATGAAGGACTGCAC---CGAGCGCCA 212  
Db 3 GlyLeuLeuGluMetTrpGln-----AspArgThrTyrHisGlyLysValPro 18  
QY 213 GGCCAACTT-----CTCCG 227  
Db 19 ArgGlnThrGlyGlyPheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38  
QY 228 CGAGGACCTGGCTTCCCCAGGGCAAGGCCCGAGTTCCCCAGCGAGCAGAA----- 281  
Db 39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerGly 58  
QY 282 CCGCGGCAACAGCCCAACAGCCGCGCA-----GCTGCAGGTGCGGGGGAACAACCCCG 335  
Db 59 ProThrGlyGluIleTyrAlaAlaArgGluLysAlaGluGlyAlaGluArgGluThrIle 78  
QY 336 CAGCGAGCGCGCGC-----CGAGCGCCA-----GGGCACCTGAACTT----- 374  
Db 79 GlnArgGlyAspArgGlyLeuAlaAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98  
QY 375 -----CCCCAGATCACCTGTGGCAGCGCCCTGGTGGAGCATCAA 416  
Db 99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValValThrAlaHi 118  
QY 417 GGTGGCGCGCGAGATCAAGAGGCGCTGTGGACACCGCGCGCGAGCACACCCCTGTCTGA 476  
Db 118 stIleGluGlyGlnProValGluValLeuLeuAspThrGlyAlaAspSerIleValAl 138  
QY 477 GAGATGAGCCTCCCGCGGAAGTGGAGCCCAAGATGATCGCGGCGATCGCGGCTTCAT 536  
Db 138 aGlyIleGluLeuGlySerAsnTyrSerProLysIleValGlyGlyIleGlyPheI 158

QY 537 CAAGGTGCGCCAGTACGACACGATCTGTATCGAGATCTCGGCAAGAGGCCATCGGCAC 596  
Db eAsnThrLysGluTyrLysAsnValGluIleGluValLeuGlyLysArgValArgAlaTh 178  
QY 597 CGTGTGTATCGCCGCCACCCCGTGAACATATCGGCGCAACATGCTGACCCAGCTGGG 656  
Db rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuG 198  
QY 657 CTGCACCTGAACTTCCCATCATCAGCCCATCAGACCGTGCCTGAACTGAAGCCCG 716  
Db yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProG 218  
QY 717 CATGACGCGCCCAAGGTGAAGCAGTGGCCCTGACCCAGGAGAGATCAAGGCCCTGAC 776  
Db yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeu 238  
QY 777 CGCCATTCGAGGAGATGAGAGAGGAGGCAAGATCACAAAGATCGGCCCGAGAACCC 836  
Db sGluIleCysGluLysMetGluArgGluGlyGlnLeuGluAlaProProThrAsnPr 258  
QY 837 CTACACACCCCGTGTTCGCCATCAAGAGAGAGACACACCACTGGCGCAAGCTGT 896  
Db oTyrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeu 278  
QY 897 CGACTTCCCGAGCTGAACAAGCGCACCCAGCACTTCTGGAGAGTGCAGCTGGGCATCCC 956  
Db eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyLePr 298  
QY 957 CCACCCCGCGCGCTGAAGAAGAAGAGCGTGCACCGTGTGGAGCGTGGCGCACCGCTA 1016  
Db oHisProAlaGlyLeuAlaLysLysArgArgIleThrValLeuAspValGlyAspAlaTy 318  
QY 1017 CTTACAGCTGCCCTCGAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGACAT 1076  
Db rPheSerIleProLeuHisGluAspPheArgGlnTyrThrAlaPheThrLeuProSerVa 338  
QY 1077 CAACACAGAGACCCCGCATCGCTTACCAGTACACAGTGTCTGCCAGCGGTGGAGGG 1136  
Db lAsnAsnAlaGluProGlyLysArgTyrIleTyrLysValLeuProGlnGlyTrpLysG 358  
QY 1137 CAGCCCGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCGCA 1196  
Db ySerProAlaIlePheGlnTyrTrpMetArgGlnValLeuGluProPheArgLysAlaAs 378  
QY 1197 CCCCAGATCGTGATC-----TACCAGGCCCTGTAGTGGCGAGCACCTGGAGAT 1250  
Db nGlnAspValIleIleIleGlnTyrMetAspAspIleLeuIleAlaSerAspArgThrAs 398  
QY 1251 CGGCCAGCACCGCGCAAGATCGAGAGCTGGCGCAAGCACCTGCTGCTGGGGCTTCAC 1310  
Db pLeuGluHisAspArgValValLeuGlnLeuLysGluLeuAsnSerLeuGlyPheSe 418  
QY 1311 CACCCCGCAGACAGACACACAGAGAGCGCCCTTCTGTGGATGGGTACAGAGTGCA 1370  
Db rThrProAspGlyLysPheGlnLysAspProProTyrArgTrpMetGlyTyrGluLeuTr 438  
QY 1371 CCCCAGACAGTGCACCGTGCAGCCCATCGAGCTGCCCGAGAGAGAGAGCTGGACCGTAA 1430  
Db pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458  
QY 1431 CGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGCGCAGCCAGATCTACCCCGCATCAA 1490  
Db nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyrProGlyIleL 478  
QY 1491 GTGCGCCAGCTGTGCAAGCTGCTGCGCGCGCAAGGCCCTTACCGACATCTGTGCCCT 1550  
Db sThrLysHisLeuCysArgIleLeuArgGlyLysMetThrLeuThrGluIleGlnTr 498  
QY 1551 GACCGAGGAGCGCGCTGAGCTGCCGAGAACCGCGAGATCCTGCGGAGCCCGTGCA 1610  
Db pThrGluLeuAlaGluAlaGluLeuGluAsnArgValIleLeuSerGlnGlnG 518

```
QY 1611 CGCGTGTACTACACCCCGACGAGGACCTGGTGGCCGAGATCCAGAAGAGGGCCACGA 1670
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
518 uGlyHisTyrTyrGlnGluGluGluGluLeuGluAlaThrValGlnLysAspGlnAspAs 538
QY 1671 CCAGTGGACCTACAGATCTACAGAGGACCTTCAAGAACCTGAAGACCGCAAGTACGC 1730
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
538 nGlnTrpThrTyrLysValHisGlnGlyGlu---LysThrLeuLysValGlyLysTyrAl 557
QY 1731 CAAGATCGCACCCGCCACCAACGACGCTGAAGCTGACCGAGGCCGTGCAGAAGAT 1790
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
557 aLysValLysAsnThrHisThrAsnGlyValArgLeuLeuAlaGlnValGlnLysIl 577
QY 1791 CGCCATCGAGAGCATCTGATCTGGGCAAGACCCCAAGTTCGCTGCCATCCAGAA 1850
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
577 eGlyLysGluAlaLeuValIleTrpGlyArgIleProLysPheHisLeuProValGluAr 597
QY 1851 GGAGACCTGGAGACCTGGTGGACCGACTACTGGCAGCGCACCTGGATCCCGAGTGGGA 1910
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
597 gGluIleTrpGluGlnTrpAspAspTyrTrpGlnValThrTrpIleProAspTrpAs 617
QY 1911 GTTCGTGAACACCCCCCTGGTGAAGCTGTGTACAGCTGGAGAAGGAGCCCATCAT 1970
Db      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
617 pPheValSerThrProProLeuValArgLeuValPheAsnLeuValLysAspProIlePr 637
QY 1971 CGCGCCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2030
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
637 oGlyAlaGluThrPheTyrThrAspGlySerCysAsnArgGlnSerLysGluGlyLysAl 657
QY 2031 CGGCTAGTCACCGACCGCGCGGAGAGATCGTGAGCGCTGACCGAGACCCACCA 2090
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
657 aGlyTyrValThrAspLysGlyArgAspLysValLysValLeuGluGlnThrThrAsnGl 677
QY 2091 GAAGACCGAGCTGCAGCCATCCAGCTGGCCCTGCAGGACAGCGGCGAGGTGAACAT 2150
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
677 nGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAspSerGlyProLysAlaAsnIl 697
QY 2151 CGTGACCGACGACGATACGCCCTGGGCATCATCCAGGCCCGCCCAAGAGCGAGAG 2210
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
697 eIleValAspSerGlnTyrValMetGlyIleValValGlyGlnProThrGluSerGluAs 717
QY 2211 CGAGCTGGTCAACAGATCATCGACGAGCTGTACAGAGGAGAGGTGTACCTGAGCTG 2270
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
717 nArgIleValAsnGlnIleGluGluMetIleLysLysGluAlaIleTyrValAlaTr 737
QY 2271 GGTGCCCGCCCAAGGGCATCGCGGCAACGAGCAGATCCACAAGCTGGTGAGCAAGGG 2330
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
737 pValProAlaHisLysGlyIleGlyGlyAsnGlnGluValAspHisLeuValSerGlnGl 757
QY 2331 CATCCGCAAGGTGCTGTTCTCGACGGCATCGAT 2364
Db      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
757 yIleArgGlnValLeuPheLeuGluLysIleGlu 768
```

Search completed: June 2, 2005, 05:08:48  
Job time : 101.833 secs



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 20:20:55 ; Search time 208.667 Seconds  
(without alignments)

12088.668 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 4573

Sequence: 1 gtcagccaccatgcccga.....gggctagcaccgtgaattc 2463

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2.1/USPTO.spool/US09610313/runat.31052005.155136.15147/app.query.fasta.1.7893  
-DB=Uniprot 03 -Qfmt=fastan -SUPFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09610313 @CGN 1.1 725 @runat.31052005.155136.15147 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Uniprot 03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	4140	90.5	1427	2	Q9WF92	Q9WF92 human immun
2	4136	90.4	1427	2	Q9WF89	Q9WF89 human immun
3	4132	90.4	1428	2	Q9WF86	Q9WF86 human immun
4	3990	87.3	1427	2	Q9WF62	Q9WF62 human immun
5	3983	87.1	1429	2	Q6X4P8	Q6X4P8 human immun
6	3955	86.5	1425	2	Q6X4Q6	Q6X4Q6 human immun
7	3948	86.3	1427	2	Q9WF96	Q9WF96 human immun
8	3944	86.2	1430	2	Q6X4R4	Q6X4R4 human immun
9	3935.5	86.1	1427	2	Q9WF71	Q9WF71 human immun
10	3931	86.0	1427	2	Q9WFA0	Q9WFA0 human immun
11	3928.5	85.9	1427	2	Q9WF65	Q9WF65 human immun
12	3927	85.9	1437	2	Q9WF77	Q9WF77 human immun
13	3919.5	85.7	1427	2	Q9WF74	Q9WF74 human immun
14	3913.5	85.6	1427	2	Q8ADX5	Q8ADX5 human immun
15	3905	85.4	1425	2	Q9WF79	Q9WF79 human immun
16	3902	85.3	1427	2	Q9WF98	Q9WF98 human immun

#### ALIGNMENTS

RESULT 1

Q9WF92	17	3893.5	85.1	1429	2	Q9WF84	Q9WF84 human immun
Q9WF92	18	3884	84.9	1428	2	Q6X4P0	Q6X4P0 human immun
Q9WF92	19	3874.5	84.7	1427	2	Q9WF68	Q9WF68 human immun
Q9WF92	20	3838.5	83.9	1432	2	Q8ADZ9	Q8ADZ9 human immun
Q9WF92	21	3835	83.9	1433	2	Q8ADE9	Q8ADE9 human immun
Q9WF92	22	3833	83.8	1434	2	Q8ADV1	Q8ADV1 human immun
Q9WF92	23	3831	83.8	1430	2	Q7SV36	Q7SV36 human immun
Q9WF92	24	3827	83.7	1425	2	Q7SV20	Q7SV20 human immun
Q9WF92	25	3821.5	83.6	1433	2	Q8ADN7	Q8ADN7 human immun
Q9WF92	26	3820.5	83.5	1433	2	Q8ADK5	Q8ADK5 human immun
Q9WF92	27	3818.5	83.5	1433	2	Q8ADS7	Q8ADS7 human immun
Q9WF92	28	3815.5	83.4	1433	2	Q8ADG5	Q8ADG5 human immun
Q9WF92	29	3812.5	83.4	1435	2	Q9DQ33	Q9DQ33 human immun
Q9WF92	30	3811.5	83.3	1433	2	Q8ADE1	Q8ADE1 human immun
Q9WF92	31	3806.5	83.2	1431	2	Q8ADR1	Q8ADR1 human immun
Q9WF92	32	3805.5	83.2	1433	2	Q8ADC5	Q8ADC5 human immun
Q9WF92	33	3804.5	83.2	1433	2	Q8AE32	Q8AE32 human immun
Q9WF92	34	3803.5	83.2	1433	2	Q8ADM1	Q8ADM1 human immun
Q9WF92	35	3803.5	83.2	1433	2	Q8ADU3	Q8ADU3 human immun
Q9WF92	36	3802.5	83.2	1437	2	O41782	O41782 human immun
Q9WF92	37	3801	83.1	1434	2	Q8Q727	Q8Q727 human immun
Q9WF92	38	3801	83.1	1435	2	Q9WLJ3	Q9WLJ3 human immun
Q9WF92	39	3800.5	83.1	1433	2	Q8AD85	Q8AD85 human immun
Q9WF92	40	3798.5	83.1	1433	2	Q7SPT9	Q7SPT9 human immun
Q9WF92	41	3794.5	83.0	1433	2	Q8ADM7	Q8ADM7 human immun
Q9WF92	42	3794.5	83.0	1436	2	Q8AD93	Q8AD93 human immun
Q9WF92	43	3792.5	82.9	1433	2	Q8AE23	Q8AE23 human immun
Q9WF92	44	3791.5	82.9	1433	2	Q8ADR9	Q8ADR9 human immun
Q9WF92	45	3789.5	82.9	1433	2	Q8ADH3	Q8ADH3 human immun

PRELIMINARY; PRT; 1427 AA.

01-NOV-1999 (TrEMBLrel. 12, Created)

01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

Gag-pol polyprotein.

Human immunodeficiency virus 1.

Viruses; Retroviral viruses; Retroviridae; Lentivirus.

NCBI\_TaxID=11676;

SEQUENCE FROM N.A.

MEDLINE=99214383; PubMed=10196340;

Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,

Foley B.T., Ndung'u T.P., Rahma M., Makhema M.J., Marlink R.,

Essex M.

"Molecular cloning and phylogenetic analysis of human immunodeficiency

virus type 1 subtype C: a set of 23 full-length clones from

Botswana."

J. Virol. 73:4427-4432(1999).

-1- SIMILARITY: Belongs to peptidase family A2.

EMBL; AF110975; AAD17135.1; -

HSSP; P04585; 1JLA.

GO: GO:0019012; C:virion; IEA.

GO: GO:0004190; P:aspartic-type endopeptidase activity; IEA.

GO: GO:0003677; F:DNA binding; IEA.

GO: GO:0008907; F:integrase activity; IEA.

GO: GO:0008233; F:peptidase activity; IEA.

GO: GO:0004523; F:ribonuclease H activity; IEA.

GO: GO:0003723; F:RNA binding; IEA.

GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.

GO: GO:0005198; F:structural molecule activity; IEA.

GO: GO:0016740; F:transferase activity; IEA.

GO: GO:0008270; F:zinc ion binding; IEA.

GO: GO:0015074; P:DNA integration; IEA.

GO: GO:0006310; P:DNA recombination; IEA.

GO: GO:0006508; P:proteolysis and peptidolysis; IEA.

GO: GO:0006278; P:RNA-dependent DNA replication; IEA.

GO: GO:0016032; P:viral life cycle; IEA.





...  
360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleuMetGlnArgSerAsnPhe 379  
73 AAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGCGCAAGAGAGGCCACATCGCCCGC 132  
380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
133 AACTGCGCGCCCCCGCGCAAGAAGGCTCTGGAAGTGGCGCAAGAGGGCCACACAGATG 192  
400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisIleMet 419  
193 AAGGACTGCACCGAGCGCCAGGCGCAACTTCTCCGCGAGGACTGCGCTTCCCGCCAGGCG 252  
420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439  
253 AAGGCGCGGAGTTCCCCAGCGAGCAGAAACCGCGCAACAGCGCCCCACCGAGCGAGCTG 312  
440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459  
313 CAGGTGCGCGCGCAGAACCCCGCAGCGAGGCGCGCGCGCGAGCGCCAGCGCCACCTGAAC 372  
460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479  
373 TTCCCCCAGATCACCTGTGGCAGCGCCCTCGTGTGAGCATCAAGGTGGCGCGCCAGATC 432  
480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle 499  
433 AAGGAGGCGCTCTGGACACCGCGCGCGCAGCACCGCTGTGGAGAGATGAGCTGCCC 492  
500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519  
493 GGCAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGCGCGCATAC 552  
520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyLysPheIleLysValArgGlnTyr 539  
553 GACCAGATCTGTATCAGATCTGCGCAAGAAGGCCATCGGCACCGTGTGATCGGCCCC 612  
540 AspGlnIleuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559  
613 ACCCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGACCCCTGAATTC 672  
560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579  
673 CCCATCAGCCCCATCAGACCGTGCCTGAGCTGAAGCCCGGCATGAGCGGCCCAAG 732  
580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
733 GTGAAGCAGTGGCCCTGACCGAGGAGAAGATCAAGCGCTGACCGCCATCTGCGAGGAG 792  
600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
793 ATGGAGAAGGAGGCAAGATCAACAAGATCGGCCCCGAGAACCCCTPACAACACCCCGTG 852  
620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
853 TTCCGCATCAAGAAGAAGACACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912  
640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
913 AACAAAGCGCACCCAGGACTTCTGGGAGGTGACGTGGGCATCCCCACCCCGCGCGCTG 972  
660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
973 AAGAAGAAGAAGCGTGCCTGTGAGCTGGCGGCGAGCCCTTCTTACGCGTGGCCCTTG 1032  
680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
1033 GACGAGGACTTCCGCAAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCCC 1092  
700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
1093 GGCATCCGCTACCAAGTACACCGTGTGCTGGCCCGAGGCTGGAAGGCGAGCCCGCATCTTC 1152

720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
1153 CAGAGCAGCATGACCAAGATCTTGAGCCCTTCCGGCGCCGCAACCCCGAGATCGTATC 1212  
740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaAsnProGluIleValIle 759  
1213 TACCAG-----GCCCCCTGTACGTGGCGCAGCGACTCGAGATCGCCAGACCCCGCC 1266  
760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
1267 AAGATCGAGGAGCTGCGCAAGCACCTGTGCTGCGCTGGGGTTTCACCAACCCCGCAAGAAG 1326  
780 LysIleGluLeuArgLysHisLeuLeuArgTyrGlyPheThrThrProAspLysLys 799  
1327 CACCAGAAGGAGCCCTTCTTCTGTGGATGGGCTAGAGCTGCACCCCGCAGCTGACC 1386  
800 HisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
1387 GTGCAGCCCATCGAGCTGCCGAGAAGAGAGCTGAGACCGTGAACCGATCCAGAAAGCTG 1446  
820 ValGlnProIleLeuLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
1447 GTGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGCATCAAGGTGCGCCAGCTGTC 1506  
840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
1507 AAGCTGCTCGCGCGCCCAAGGCCCTGACCGACATCGTCCCTGACCGAGGCGCGAG 1566  
860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
1567 CTGGAGCTGCGCAGAACCGCAGATCTCTGCGAGCCCGTGCACCGCTGACTACTACGAC 1626  
880 LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTyrAsp 899  
1627 CCCAGCAAGAGACTGTGTGGCCGAGATCCAGAAGCGGCGCACGACAGTGGACTTACCAG 1686  
900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
1687 ATCTACGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCAGCGCC 1746  
920 IleTyrGlnGluProPheLysAsnLeuTyrGlyLysTyrAlaLysMetArgThrAla 939  
1747 CACACCAACGAGCTGAAGCAGCTGACCGAGGCGCTGCAGAAGATCCCATCGAGAGCATC 1806  
940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
1807 GTGATCTGGGCAAGACCCCAAGTTCCCGCTGCCATCCAGAAGAGACCTGGGAGACC 1866  
960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
1867 TGGTGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCC 1926  
980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
1927 CCCTCGTCAAGCTGTGTACCAAGCTGGAGAGGACCCCATCATCGCGCGCGAGACCTTC 1986  
1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
1987 TACGTGGAGCGCGCCCAACCGCAGACCAAGATCGGAAGCGCCGGCTACGTGACCGAC 2046  
1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
2047 CGGGCGCGCAGAGATCTGTAGCTTGACCGAGACCCACCAACGAGAGACCGAGCTGCAG 2106  
1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
2107 GCATCCAGCTGGCCCTCGAGGACAGCGCAGCGAGGTGAACATCGTGCAGCAGCCAG 2166  
1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
2167 TAGCCCTCGGCATCATCCAGCGCCAGCCGACAAAGAGCGAGAGCGAGCTGTGTGAACAG 2226  
1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099



Db 580 ProileSerProileGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
QY 733 GTCAACAGTGGCCCTGACCGAGGAGAGATCAAGCCCTGACCCCTATCTCGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysLeuLysAlaLeuThrAlaLeuLysGluGlu 619  
QY 793 ATGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 852  
Db 620 MetGluLysGluGlyLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 639  
QY 853 TTCGCCATCAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 912  
Db 640 PheAlaLeuLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
QY 913 ACACAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyLeuProAlaGlyLeu 679  
QY 973 AGAAGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu 699  
QY 1033 GACGAGGACTTCCGGAAGTACACCGCTTCCATCCCGAGCATCAACACGAGACCCCC 1092  
Db 700 AspGluAspPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GGCATCCGCTACAGTACAGCTGCTGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1152  
Db 720 GlyIleArgTrpGlnTrpAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
QY 1153 CAGAGCAGATGACCAAGATCTCTGGAGCCCTTCCGGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysLysLeuGluProPheArgAlaArgAsnProGluLeuValle 759  
QY 1213 TACCAG-----GCCCCCTGTAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
QY 1267 AGATCGAGGAGTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1326  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrProAspLysLys 799  
QY 1327 CACCAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1386  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
QY 1387 GTGACGCCATCAGTGTGCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1446  
Db 820 ValGlnProileGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
QY 1447 GTGGGCAAGTGAAGTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1506  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuLys 859  
QY 1507 AAGCTGTGTGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1566  
Db 860 LysLeuLeuArgLysAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1567 CTGAGCTGGCGGAGAACCGGAGATCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1626  
Db 880 LeuGluLeuAlaGluAspArgGluIleLeuLeuArgGluProValHisGlyValTyrTrpAsp 899  
QY 1627 CCCAGCAGGAGCTGTGGCGGAGATCCAGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1686  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTrpGln 919  
QY 1687 ATCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1746  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTrpAlaLysMetArgThrAla 939  
QY 1747 CACACCAACGAGCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1806  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

QY 1807 GTGATCTGGGCAAGAGACCCCAAGTTCCGCTGCCCATCCAGAAAGAGACCTGGGAGACC 1866  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1867 TGGTGACCGACTACTGGCAGGCCACTCGATCCCGAGTCCCGAGTGGAGTTCGTGAACACCCCC 1926  
Db 980 TrpTrpTrpAspTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY 1927 CCCCTGGTGAAGTGTGTACAGCTGGAGAGGAGGCCCATCATCCGCGCCGAGACCTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleGlyAlaGluThrPhe 1019  
QY 1987 TACGTGGAGCGCGCCCAACCGCAGAGCAAGATCGGCAAGCGCCGCTACGTGACCCGAC 2046  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
QY 2047 CGGGGCGCGCAGAGATCTGTGACCTGACCGAGACCAACAGAGACCCGAGGTCGAG 2106  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
QY 2107 GCATCCAGCTGGCCTGCAGGACGCGCAGGAGGTGAACATCGTGACCCGACGACGAG 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
QY 2167 TACGCTCGGATCATCTCAGGCCCGCAGCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 2226  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluLeuValAsnGln 1099  
QY 2227 ATCATCGACAGCTGTATCAAGAGGAGAGGTGTACTCTGAGCTGGTGGTCCCGCCCAAG 2286  
Db 1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2287 GGCATCGGCGGCAACGAGCAGATCGACAGCTGTGAGCAAGGAGGAGGAGGAGGAGGAGGAG 2346  
Db 1120 GlyIleGlyLysAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139  
QY 2347 TTCCTGCGCGGATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 4  
Q9WF62 PRELIMINARY; PRT; 1427 AA.  
AC Q9WF62; (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.,  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432 (1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110961; AAD17021.1; -.  
DR HSSP; Q70622; IHVN.  
DR GO; GO:0019012; C:viroin; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA directed DNA polymerase activity; IEA.





Qy	1387	GTGCAGCCCATCGAGCTGCCCGAAGAGGAGAGCTGGACCGCTGAACGACATCCAGAAGCTG	1446
Db	820	ValGlnProIleAsnLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu	839
Qy	1447	GTGGCGAGCTGAAGTGGCGCCAGCCAGACTTACCCCGGCATCAGGTGGCCGAGCTGTGC	1506
Db	840	ValGlyLysLeuAsnTrpAlaSerGlnIleTrpProGlyLysValArgGlnLeuCys	859
Qy	1507	AAGCTGTGTCGCGCGCCCAAGGCCCTGCACCGACATCTGCGCCCTGCACCGAGGAGGCCGAG	1566
Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu	879
Qy	1567	CTGAGCTGGCCGGAACCCGCGAGATCCTGCGCGAGCCCGCTGCACGCGCTGTACTACGAC	1626
Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
Qy	1627	CCGAGCAGGAGCTGGTGGCGGAGATCCAGAGCAGGCGCCACGACAGTGGAGCTACCAG	1686
Db	900	ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln	919
Qy	1687	ATCTACCAAGAGCCCTTCAAGAACCTCGAAGCAGCGCAAGTACGCCAAGATGCGCACCGCC	1746
Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
Qy	1747	CACACCAACGAGCTGAAGCAGCTCACCGAGCCCTGCAGAGAATCGCCATCGAGAGCATC	1806
Db	940	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle	959
Qy	1807	GTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAAAGGAGAGACTGGAGACC	1866
Db	960	ValIleTrpGlyLysTrpProLysPheArgLeuProIleGlnLysGluThrTrpGluThr	979
Qy	1867	TGGTGGACCGACTACTCGCGCCACCTGGATCCCGAGTCCCGAGTGGAGTTCGTGAACACCC	1926
Db	980	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
Qy	1927	CCCCTGTGTGAAGCTGTGGTACCACTCGAGAGAGAGCCCATCATCTGCGCGCGAGACCTTC	1986
Db	1000	ProLeuValLysLeuTrpTyrGlnLeuGluGluProIleValGlyAlaGluThrPhe	1019
Qy	1987	TACGTGACGGCGCGCCCAACCCGCGAGACCAAGATCGGCAAGGCCGCTACGTGACCGAC	2046
Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp	1039
Qy	2047	CGGGCGCGCAGAGATCGTGAGCCTCACCGAGACCAACCAAGAGAGACCGAGCTGCAG	2106
Db	1040	ArgGlyArgGlnArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln	1059
Qy	2107	GCCATTCAGCTGGCCCTCGCAGCAGCGGAGCGGAGTGAACCTGTAACCTGACCGACCGAG	2166
Db	1060	AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln	1079
Qy	2167	TACGCCCTGGGCATCATCCAGGCCCGCCGACAAAGACGAGACCGAGCTGGTGAACAG	2226
Db	1080	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln	1099
Qy	2227	ATCATCAGACGCTGATCAAGAAGAGAGGTGTACCTGAGCTGGGTGCCGCCACCAAG	2286
Db	1100	IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys	1119
Qy	2287	GGCATCGCGGCAACGACGAGATCGACAAGCTGGTGAGCAAGGCGCATCCCGAGGTGCTG	2346
Db	1120	GlyIleGlyLysGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu	1139
Qy	2347	TTCCTGGACGGCATCGAT	2364
Db	1140	PheLeuAspGlyIleAsp	1145
RESULT 5			
Q6X4P8			
ID Q6X4P8			
AC Q6X4P8:			
		PRELIMINARY;	PRT; 1429 AA.

## RESULTS

06X4P8

ID O6X4P8

AC Q6X4P8;

PRELIMINARY: PRT: 1429 AA.

[illegible]

05-JUL-2004 (TrEMBLrel. 27, Created)  
05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
Gag-pol fusion polypeptide (Fragment).  
Human immunodeficiency virus 1.  
Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
[1]  
SEQUENCE FROM N.A.  
PubMed=14709249; DOI=10.1089/08922203771881220;  
Harris M.E., Birx D.L.;  
Maayan S., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.  
McCutchan F.E.;  
"A cluster of HIV type 1 subtype C sequences from Ethiopia,  
in full genome analysis, is not sustained in subgenomic regi  
AIDS Res. Hum. Retroviruses 19:1125-1133(2003).  
[2]  
SEQUENCE FROM N.A.  
Harris M.E., Birx D.L.;  
Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A.  
Kim B., McCutchan F.E.;  
Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
[4]  
SEQUENCE FROM N.A.  
Maayan S., Zeira M.;  
Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
-1- SIMILARITY: Belongs to peptidase family A2.  
EMBL; AY255825; AAP76538.1; -.  
HSP; P03367; IABG.  
GO; GO:0001902; C:viralon; IEA.  
GO; GO:0004190; P:aspartic-type endopeptidase activity; IEA.  
GO; GO:0003677; P:DNA binding; IEA.  
GO; GO:0008907; P:integrase activity; IEA.  
GO; GO:0008223; P:peptidase activity; IEA.  
GO; GO:0004523; P:ribonuclease H activity; IEA.  
GO; GO:0003723; P:RNA binding; IEA.  
GO; GO:0003964; P:RNA-directed DNA polymerase activity; IEA.  
GO; GO:0005198; P:structural molecule activity; IEA.  
GO; GO:0016740; P:transferase activity; IEA.  
GO; GO:0008270; P:zinc ion binding; IEA.  
GO; GO:0015074; P:DNA integration; IEA.  
GO; GO:0006310; P:DNA recombination; IEA.  
GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
GO; GO:0016032; P:viral life cycle; IEA.  
InterPro; IPR000721; Gag\_p24.  
InterPro; IPR001037; Integrase C.  
InterPro; IPR003308; Integrase Zn\_N.  
InterPro; IPR001995; Peptidase A2.  
InterPro; IPR009007; Pept Aspartic.  
InterPro; IPR001969; Pept\_Asp\_AS.  
InterPro; IPR010999; Retrovir\_matrix.  
InterPro; IPR000071; Retrovir\_p17.  
InterPro; IPR008916; Retrov capsid C.  
InterPro; IPR008919; Retrov capsid\_N.  
InterPro; IPR002156; RNaseH.  
InterPro; IPR001584; RVT.  
InterPro; IPR000477; RVTse.  
InterPro; IPR010659; RVT\_connect.  
InterPro; IPR010661; RVT\_thumb.  
InterPro; IPR001878; Znf\_CCHC.  
Pfam; PF00540; Gag\_p17; 1.  
Pfam; PF00607; Gag\_p24; 1.  
Pfam; PF02022; Integrase\_Zn; 1.  
Pfam; PF00075; RNaseH; 1.  
Pfam; PF00665; rve; 1.  
Pfam; PF00077; RVP; 1.  
Pfam; PF00078; RVT\_1; 1.  
Pfam; PF08815; RVT\_Connect; 1.  
Pfam; PF06817; RVT\_Chumb; 1.  
Pfam; PF00098; zf\_CCHC; 2.  
PRINTS; PR00939; C2HCZFINGER.

L., observed  
ons.";

DR PRINTS; PR00234; HIV1MATRIX.  
 DR SMART; SM00343; ZNF C2HC; 2.  
 DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS50175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS50158; ZF CCHC; 2.  
 KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferrase; Zinc;  
 KW Zinc-finger.  
 FT NON\_TER 1  
 SO SEQUENCE 1429 AA; 161586 MW; DOB9CD124DC64E19 CRC64;

Alignment Scores:  
 Pred. No.: 6.56e-150 Length: 1429  
 Score: 3983.00 Matches: 743  
 Percent Similarity: 97.47% Conservative: 27  
 Best Local Similarity: 94.05% Mismatches: 14  
 Query Match: 87.10% Indels: 6  
 DB: 2 Gaps: 2

US-09-610-313B-31 (1-2463) x Q6X4P8 (1-1429)

QY	13	ATGCCGAGGCGCATGAGCGAGCCAGCCAGCCCAACATCTCTGATCGAGCGCAGCAACTTC	72
DB	358	LeuAlaGluAlaMetSerGlnAlaAsnAlaAsnMetMetGlnLysSerAsnPhe	377
QY	73	AAGGCGCCCAAGCCATCATCAAGTCTTCACTGCGGCAAGAGGCGCCATCGCCCGC	132
DB	378	LysGlyProLysArgIleLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg	397
QY	133	AACCTGCGCGCCCGCCAGAGGGCTGCTGGAGTGCGCGCAAGAGGGCCACCATG	192
DB	398	AsnCysArgAlaProArgLysArgLysCysTrpLysCysGlyLysGluGlyHisGlnMet	417
QY	193	AAGGACTGCCAGCGAGCCAGCCCAACTTCTTCGCGAGGACCTTGGCTTCCCCCAGGGC	252
DB	418	LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheGlnGly	437
QY	253	AAGGCGCGAGTTCCTCCAGGAGCAGAAACCGCCCAAGAGCCCAACAGCGCGAGCTG	312
DB	438	GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu	457
QY	313	CAGTGGCGCGGCAACACCCCGAGCGAGCGCGCGCCCGCCAGCGCCAGGGC-----	363
DB	458	GlnValArgGlyAspAsnProLysSerGluAlaGlyAspGluArgGlnGlyAlaLeuGln	477
QY	364	---ACCTGAACTTCCCGCCAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTG	420
DB	478	ArgThrLeuAsnCysProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIle	497
QY	421	GGCGCCAGATCAAGGAGCGCTGCTGGACACCGCGCCCGACGACACCGTCTGGAGGAG	480
DB	498	GlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGlu	517
QY	481	ATGAGCTCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAG	540
DB	518	IleAsnLeuProGlyLysTrpLysProLysMetIleGlyLysIleGlyLysPheIleLys	537
QY	541	GTGCGCCAGTACACCAAGATCTCTGATCGAGATCTGCGCGCAAGAGGCGCATCGGCGCGT	600
DB	538	ValArgGlnTrpAspGlnIleLeuIleGluLysCysGlyLysAlaIleGlyThrVal	557
QY	601	CTGATCGGCGCCCGCCGCGGAGCATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGC	660
DB	558	LeuValGlyProThrProValAsnIleGlyArgAsnMetLeuThrGlnLeuGlyCys	577
QY	661	ACCTGAACTTCCCATCGCCCATCGACACCGTGGCGCGCGTGAAGCTGAAGCGCGCATG	720
DB	578	ThrLeuAsnPheProLysSerProIleGluThrValProValLysLeuLysProGlyMet	597
QY	721	GACGGCGCCCAAGTGAAGCAGTGGCCCTCACCGAGGAGAGATCAAGGCCCTGACCGCC	780
DB	598	AspGlyProLysValLysGlnTrpProLeuThrLysGluLysIleGluAlaLeuThrAla	617

QY	781	ATCTCGGAGGAGATGGAGAAGAGGCGCAAGATCACCAAGATCGGCGCCGAGAACCCCTAC	840
DB	618	IleCysGluGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyr	637
QY	841	AACACCCCGTGTTCGCATCAAGAGAAGACAGACCAAGTGGCGCAAGCTGTGGAC	900
DB	638	AsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAsp	657
QY	901	TTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGATCCCCAC	960
DB	658	PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyLeuProHis	677
QY	961	CCCGCGCGCTGAAGAAGAAGAGCGTGACCGTGTGGAGCTGGCGAGCGCTACTTC	1020
DB	678	ProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe	697
QY	1021	AGCGTGGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAAC	1080
DB	698	SerValProLeuAspGluGlyPheArgLysTyrThrAlaPheThrIleProSerIleAsn	717
QY	1081	AACGAGACCCCGCGCATCCGCTACAGTACAACTGCTGCCCGCGGCTGGAAGGGCAGC	1140
DB	718	AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer	737
QY	1141	CCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGGAGCCCTTCCGCGCGCCCAACCC	1200
DB	738	ProAlaIlePheGlnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnPro	757
QY	1201	GAGATCGTGATCTACACG-----GCCCGCTGTACGTGGCGAGCGACTGGAGTCGC	1254
DB	758	GluLeuValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGly	777
QY	1255	CAGCACCGCCCAAGATCGAGGAGCTCGCAAGCACCTGCTGCGCTGGGCTTCCACACC	1314
DB	778	GlnHisArgAlaLysIleGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThr	797
QY	1315	CCGCAAGAAGACACCAAGAGAGCGCCCTTCTCTGTGGATGGGTACGAGCTGCACCC	1374
DB	798	ProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisPro	817
QY	1375	GACAAGTGGACCTGCGAGCCATCGAGCTCGCGAGGAGGAGAGCTGGACGTCGCAACGAC	1434
DB	818	AspLysTrpThrValGlnProIleGlnLeuProGluLysGluSerTrpThrValAsnAsp	837
QY	1435	ATCCAGAAGCTGTGGCAAGCTGAATGGCGCCAGCCAGATCTACCCCGGCATCAAGTG	1494
DB	838	IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysVal	857
QY	1495	CGCGAGCTGTGCAAGCTGTGCGCGCGCCAAAGCCCTGACCGACATCTGTGCCCTGACC	1554
DB	858	ArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThr	877
QY	1555	GAGGAGCGCGAGCTGGAGCTGGCGGAGAACCGGAGATCTCTGCGGAGCGCGTGCACGC	1614
DB	878	GluGluAlaGluLeuGluLeuAlaGluAsnArgGlnLeuLysGluProValHisGly	897
QY	1615	GTGTACTACACCCCGAGAGGACCTGTGCGCGAGATCCAGAGATCCAGAGAGGCGCACGAC	1674
DB	898	ValTyrTyrAspProSerLysAspLeuIleAlaGluLeuGlnLysGlnGlyHisAspGln	917
QY	1675	TGACCTTACAGATCTACCGAGGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG	1734
DB	918	TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys	937
QY	1735	ATCGGACCGCCACACCAACGAGCTGAGCAGCTGACCGAGCGCGTGCAGAGATCGCC	1794
DB	938	MetArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla	957
QY	1795	ATGAGAGCATCTGTGATCTGGGCAAGACCCCAAGTTCGCTGCCCTCATCCAGAGGAG	1854
DB	958	MetGluGlyIleValIleTrpGlyLysValProLysPheArgLeuProIleGlnLysGlu	977
QY	1855	ACCTGGGAGACCTGTGTGGACCGGCGGCTGTGGAGCCACCTGGATCCCGAGTGGAGTTC	1914





RESULT 7  
Q9WF96 PRELIMINARY; PRT; 1427 AA.  
AC Q9WF96;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makheima M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
CC -I- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110976; AAD17144.1; -;  
DR HSP; Q70622; 1HN.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR001037; Integrase\_C.  
DR InterPro; IPR003308; Integrase\_N.  
DR InterPro; IPR001995; Peptidase\_A2.  
DR InterPro; IPR009007; Pept\_Aspartic.  
DR InterPro; IPR001969; Pept\_Asp\_AS.  
DR InterPro; IPR000071; Retrovir\_matrix.  
DR InterPro; IPR008916; Retrovir\_p17.  
DR InterPro; IPR008919; Retrov\_capsid\_C.  
DR InterPro; IPR008919; Retrov\_capsid\_N.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVTse.  
DR InterPro; IPR010659; RVT\_connect.  
DR InterPro; IPR010661; RVT\_thumb.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF02022; Integrase\_Zn; 1.  
DR Pfam; PF00075; RNaseH\_1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT\_1; 1.  
DR Pfam; PF06815; RVT\_connect; 1.  
DR Pfam; PF06817; RVT\_thumb; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.  
DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
DR PROSITE; PS01175; ASP\_PROT\_RETROV; 1.  
DR PROSITE; PS0158; ZF\_CCHC; 2.

KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
KW Zinc-finger. 1427 AA; 160518 MW; FBDA73D603A68B18 CRC64;  
SQ SEQUENCE

## Alignment Scores:

Pred. No.:	1-59e-148	Length:	1427
Score:	3948.00	Matches:	730
Percent Similarity:	97.07%	Conservative:	33
Best Local Similarity:	92.88%	Mismatches:	21
Query Match:	86.33%	Indels:	2
DB:	2	Gaps:	1

US-09-610-313B-31 (1-2463) x Q9WF96 (1-1427)

QY	13	ATGGCGGAGGCGCATGAGCCAGCGCCACCGCAACATCTCTGATGAGCGGCGAGCACTTC	72
Db	360	LeuAlaGluAlaMetSerGlnAlaAsnAenThrAsnIleMetIleGlnArgSerAenPhe	379
QY	73	AAGGGCCCCAAGCGCATCATCAAGTGCTTCAACTGGCGCAAGGAGGGCCACATCGCCCGC	132
Db	380	LysGlyProArgSerValLysCysPheAsnCysGlyGluGlyHisIleAlaArg	399
QY	133	AACTGCGCGCGCCCGCGCAAGAGGGCTCTCGAAGTGGCGCAAGAGGGCCACCAAGATG	192
Db	400	AsnCysArgAlaProArgLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet	419
QY	193	AAGGATGTCACGAGCGCCAGCGCCAACTTCTTCGGGAGGAGACTGGCTTCCCCAGGGC	252
Db	420	LysAspCysThrGluArgGlnAlaAsnPheGlyGluAsnLeuAlaPheProGlnGly	439
QY	253	AAGGCGCGGAGTTCCTCCAGCGGAGGAGACCGCGCCACAGCCACAGCGCGGAGCTG	312
Db	440	GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu	459
QY	313	CAGGTGCGCGCGCACAAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	372
Db	460	ArgValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn	479
QY	373	TTCCCCCAGATCACCTGTGGGAGCGCCCGCTGTGGTGGAGCATCAAGTGGCGCGCGC	432
Db	480	PheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysValGlyGlnVal	499
QY	433	AAGGAGCGCTCTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	492
Db	500	LysGluAlaLeuLeuAspThrGlyAlaAspThrValLeuGluGluLeuLeuLeuPro	519
QY	493	GGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGCATC	552
Db	520	GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr	539
QY	553	GACCAGATCTGATCGAGATCTCGCGCAAGAGGCCATCGGACCGTGTGATCGCGCGCC	612
Db	540	AspGlnIleThrIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro	559
QY	613	ACCCCGTGAACATCATCGCGCGCAACATGTCACCGCTGGCTGGCTGACCTGAACCTTC	672
Db	560	ThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe	579
QY	673	CCCATCAGCGCCCGCTGAGACCGGTGCGCGTGAAGCTGAAGCCCGGCGCATGAGCGCC	732
Db	580	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
QY	733	GTGAGCAGTGGCGCTGACCGGAGGAGATCAAGCGCCCTGACCGCCCATCTCGCGAGG	792
Db	600	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu	619
QY	793	ATGGAGAAGGAGGCAAGATCACCAAGATCGCGCGCGAGAACCCCTACAAACACCCCGT	852
Db	620	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProLysThrAsnThrPro	639
QY	853	TTTCGCCATCAAGAAGAGGAGCACCAAGTGGCGCGCAAGCTGGTGGACTTCGCGGAGCT	912

Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAGCGCCACCCAGGACTTCTGGAGAGTTCAGCTGGGCATCCCCACCCCGCGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAGAGAGCGTGCCTGCTGACCTGGCGACGCTACTTTCAGCGTCCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu 699  
Qy 1033 GACGAGGACTTCCCAAGTACACCGCTTACCATCCCGACATCAACAACGAGACCC 1092  
Db 700 AspGluGlyPheArgLysTrpThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719  
Qy 1093 GGATCCGCTACCACTACAACGCTGCTGCCAGGGCTGGAAGGACGCCACGATCTTC 1152  
Db 720 GlyIleArgTrpGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy 1153 CAGAGCATGACCAAGATCCTGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759  
Qy 1213 TACCAG-----GCCCCCTCTAGTGGGACGACCTGGAGATCGGCGCACCGCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCCACCCCGCGCAAGAAG 1326  
Db 780 LysIleGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799  
Qy 1327 CACCAAGAGAGCCCCCTCTCTGTGATGGGTACAGAGTGCACCCCGCAAGTGGACC 1386  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1387 GTGCAGCCCTCAGAGTCCCGCAAGAGAGAGTGGACCGTGAACGACATCCCAAGCTG 1446  
Db 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1447 GTGGCAAGCTGAACTGGGCGCAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTC 1506  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
Qy 1507 AAGTCTGTCGGCGCGCAAGGCCCTGACGACATCTGTGCCCTGACCGAGGAGCGCAG 1566  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879  
Qy 1567 CTGAGCTGCGCAGAACCGAGATCTCGCGAGCCCGTGCACGCGCTACTACGAC 1626  
Db 880 LeuGluLeuAlaGluAsnAsgluIleLysGluProValHisGlyValTyrTrpAsp 899  
Qy 1627 CCCAAGAGGACCTGTGGCGGAGATCCAGAGCAGGCGCCACGACCATGAGCTACACG 1686  
Db 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1687 ATTCACGAGGACCTTCAGAACTGAGACCGCGCAAGTACGCGCAAGATGCGCACCGCC 1746  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1747 CACACCAACCGTGAAGAGCTGACCGAGCGCTGCAGAGATCGCCATGGAGAGCATC 1806  
Db 940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959  
Qy 1807 GTGATCTGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGAGACC 1866  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1867 TGTGGACCGACTACTGGCAGGCCACTTGGATCCCGAGTGGGAGTTCTGTAACACCC 1926  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1927 CCGCTGTGAGCTGTGTACCGAGAGAGAGAGCCCATCATCGCGCGCGAGACCTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019

Qy 1987 TACCTGACCGCGCGCCGCAACCCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCCAG 2046  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2047 CGGGCGCGCAGAGATCGTGCAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAG 2106  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
Qy 2107 GCATCCAGCTGGCCCTGCGAGGACGCGGAGGAGGAGTGAACATCGTGACCGACGAC 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079  
Qy 2167 TACGCTCGGCGCATCTCCAGGCCCGCCGACCAAGAGCGAGAGCGAGCTGTTGAACCCAG 2226  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
Qy 2227 ATCATCCAGCAGCTGATCAAGAGGAGAGAGTGTACCTGAGCTGGTGGCGCCGACCAAG 2286  
Db 1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy 2287 GGCATCCGCGCGCACGAGCAGATCCACAAGCTGTGAGCAAGGCGCATCCGCAAGTGTG 2346  
Db 1120 GlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy 2347 TTCCTGACCGCATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 8  
Q6X4R4 ID Q6X4R4 PRELIMINARY; PRT; 1430 AA.  
AC Q6X4R4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DE Gsg-poi fusion polyprotein (Fragment).  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14709249; DOI=10.1089/08922203771891220;  
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birx D.L.,  
RA McCutchan F.E.;  
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
RT in full genome analysis, is not sustained in subgenomic regions.";  
RL AIDS Res. Hum. Retroviruses 19:1125-1133 (2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Harris M.E., Birx D.L.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kim B., McCutchan F.E.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Maayan S., Zeira M.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR HSPB; AY255823; AAF76521.1; -;  
DR HSPB; P12497; I89D.  
DR GO; GO:0019012; C:viron; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:protease activity; IEA.  
DR GO; GO:0008233; F:protease activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.





```
QY 1375 GACAAGTGGACCTGGACCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGAC 1434
Db 819 AsplYsTrpThrValGlnThrIleGlnLeuProGluLysAspSerTrpThrValAsnAsp 838
QY 1435 ATCCAGAAGCTGGTGGCAAGCTGAACCTGGCGCCAGCCAGATCTACCCCGGCATCAAGGTG 1494
Db 839 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysLeuVal 858
QY 1495 CGCAGCTGTGCAAGCTGTGGCGCGGCCCAAGCCCTGACCGACATCTGTGCGCCCTGACC 1554
Db 859 ArgGlnLeuGlyLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysLeuVal 878
QY 1555 GAGAGGCCAGCTGGAGCTGGCGCGGCCCAAGCCCTGACCGACATCTGTGCGCCCTGACC 1614
Db 879 GluGluAlaGluLeuGluAlaGluAsnArgGlnIleLeuLysGluProValHisGly 898
QY 1615 GTGTACTTACCAACCCAGACGAGCTGGTGGCGCGGCCCAAGCCCTGACCGACATCTGTGCGCCCTGACC 1674
Db 899 ValTyrTrpAspProSerLysAspLeuIleAlaGluLeuGlnLysGlnGlyHisAspGln 918
QY 1675 TGGACCTTACCAAGTCTTACCAAGGAGCCCTTCAAGAACCTGGAAGCCCGCAAGTACGCCAAG 1734
Db 919 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 938
QY 1735 ATCGGACCCGCCACCAAGGAGCTGAAGCAGCTGACCGAGCCCGTGGCAGAGATCGCC 1794
Db 939 ArgArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 958
QY 1795 ATGGAGAGCTGTGATCTCGGCGCCAGACCCCAAGTTCGCGCTGCCATCCAGAGGAG 1854
Db 959 MetGluSerIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGlu 978
QY 1855 ACCTGGGAGACCTGGTGGACCGCACTACTGGCAGCGCCACCTGGATCCCGGAGTGGAGTTC 1914
Db 979 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 998
QY 1915 GTGAACACCCCGCTGGTGAAGCTGTGTGTAAGCTGGAGAGGAGCCCATCATCGGC 1974
Db 999 ValAsnThrProProLeuValLysLeuTyrTrpGlnLeuGluLysGluProIleValGly 1018
QY 1975 GCCGAGACCTTCTACGTGGAGCGCGCGCCCAACCGCGAGACCAAGATCCGCGAGCGCGC 2034
Db 1019 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1038
QY 2035 TACGTGACCCAGCGCGCGCGCAAGATCGTGAAGCTGACCGAGACCAAGACCAAGCAAG 2094
Db 1039 TyrValThrAspLysGlyArgGlnLysValSerLeuAsnGluThrThrAsnGlnLys 1058
QY 2095 ACCGAGTGCAGGCCATCCAGCTGGCCCTGCGAGACAGCGCGAGCGAGTGAACATCGTG 2154
Db 1059 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1078
QY 2155 ACCGAGACCGAGTACGCCCTGGGATCATCCAGCCCGAGCCCGAGAGCGAGAGCGAG 2214
Db 1079 ThrAspSerGlnTyrAlaLeuGlyIleGlnAlaGlnProAspLysSerAspSerGlu 1098
QY 2215 CTGCTGAACAGATCATCGAGAGCTGATCAAGAAGGAGAGTGTACCTGAGCTGGGTG 2274
Db 1099 IleValAsnGlnIleIleGlnLeuIleLysLysGluArgValTyrLeuSerTrpVal 1118
QY 2275 CCCGCCACCAAGGCGATCGCGCGCAAGCAGCAGATCGACAGCTGTGTGAGCAAGGGCATC 2334
Db 1119 ProAlaHisLysGlyIleGlyLysAsnGluValAspLysLeuValSerLysGlyIle 1138
QY 2335 CGCAAGTGTCTTCTGGACCGCATCGAT 2364
Db 1139 ArgLysValLeuPheLeuAspGlyIleAsp 1148
RESULT 9
Q9WF71 PRELIMINARY; PRT; 1427 AA.
AC Q9WF71;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
```

```
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.;
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana.";
RL J. Virol. 73:4427-4432(1999).
CC -1- SIMILARITY: Belongs to peptidase family A2.
DR EMBL; AF110964; AAD17047.1; --
DR HSP; Q70622; IHN.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn_N.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR009007; Pept_Asp_AS.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010659; RVT_connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF005540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIV1MATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR PROSITE; PS00158; ZF_CCHC; 2.
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polypeptide; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 161326 MW; B9FC43C29B26CFA2 CRC64;
```

Alignment Scores:

Pred. No.:	4,96e-148	Length:	1427
Score:	3935.50	Matches:	727
Percent Similarity:	96.95%	Conservative:	35
Best Local Similarity:	92.49%	Mismatches:	21
Query Match:	86.06%	Indels:	3
DB:	2	Gaps:	2
US-09-610-313B-31 (1-2463) x Q9WP71 (1-1427)			
QY	13	ATGGCGAGGCGCATGAGCGAGCCAGCGCCCAACATCTGATGCGAGCGCAGCACTTC	72
Db	361	LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe	379
QY	73	AAGGGCCCCAAGCGCATCATCAAGTCTTCAACTGGCGCAAGAGAGGCCACATCGCCCGC	132
Db	380	LysGlyProArgArgileValLysCysPheAsnCysGlyLysGluGlyHisleAlaArg	399
QY	133	AACTGCGCGCGCCCCCGCAAGAAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACCAAGATG	192
Db	400	AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet	419
QY	193	AAGGACTGCACGAGCGCCAGGCCCAACTTCTTCCGCGAGGACTGGCCCTTCCCCAGGGC	252
Db	420	LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly	439
QY	253	AAGGCCCGCGAGTTCCCCAGCGAGCAGAAACCGCGCCCAACAGCGCCCAACAGCGCGCAGCTG	312
Db	440	GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu	459
QY	313	CAGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	372
Db	460	GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyAsnPheAsn	479
QY	373	TTCCCCCAGATCACCTGTGGCAGCGCCCGCCCTGGTGAGCATCAAGTGGCGCGCGCAGATC	432
Db	480	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle	499
QY	433	AAGGAGGCCCTGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	492
Db	500	ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro	519
QY	493	GGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGCAGTAC	552
Db	520	GlyLysTrpLysProLysIleIleGlyLysIleGlyPheIleLysValArgGlnTyr	539
QY	553	GACCATCTGATTCAGATCTCGCGCAAGAGGCCATCGCACCTGGACCTGTGATCGCGCCC	612
Db	540	GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro	559
QY	613	ACCCCGCTGAACATCATCGCGCGCAACATGCTGACCGCAGCTGGCGCTGCACCCCTGAACCTTC	672
Db	560	ThrProIleAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe	579
QY	673	CCCATCAGCCCCCATCGAGACCGTGGCCGCTGAAGCTGAAGCCCGCGCATGACCGCCCCAAG	732
Db	580	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
QY	733	GTGAAGCAGTGGCCCTGTGACCGAGGAGAGATCAAGGCCCTGACCCGCTATCTGCGAGGAG	792
Db	600	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu	619
QY	793	ATGGAGAAGGGGGAAGATCACCAAGATCGCGCGCGCGAGAACCCCTACCAACACCCCGCTG	852
Db	620	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTrpAsnThrProVal	639
QY	853	TTTCGCCATCAAGAAGGACAGCACCAAGTGGCGCAAGCTGGTGAGCTTCCGCGAGCTG	912
Db	640	PheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
QY	913	ACAAGCGCACCCAGACTTCTGGAGGTGCGAGCTGGGCATCCCCACCCCGCGCGCGCTG	972
Db	660	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679

QY	973	AAGAAGAAGAGAGCGGTGACCGTGTGGACGTGGGCGAGCGCTACTTCAAGCTGCCCTTG	1032
Db	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
QY	1033	GACGAGGACTTCCGCAAGTACACCGCCTTCAACATCCCGCAGCATCAACACGAGACCCCC	1092
Db	700	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro	719
QY	1093	GCATCCCGCTACAGTACAAACGTGTGCTGCCAGGGCTGGAAGGCGAGCCCGCAGCATCTTC	1152
Db	720	GlyIleArgTyrGlnTrpAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	739
QY	1153	CAGACGAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC	1212
Db	740	GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
QY	1213	TTACCAG-----GCCCCCTGTACGTGGCGACGACCTGGAGATCGGCGCAGCACCGCGCC	1266
Db	760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
QY	1267	AAGATCGAGGAGCTGCGCAAGCACCTGTGCTGGGGCTTCAACACCCCGCAGCAAGAAG	1326
Db	780	LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys	799
QY	1327	CACCAAGAGGAGCCCTTCTCTGTGGATGGCTACGAGCTGCACCCCGCAAGTGGACC	1386
Db	800	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
QY	1387	GTGAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTG	1446
Db	820	ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu	839
QY	1447	GTGGGCAAGCTCAACTGGGCGCAGCAGATCTACCCGGCATCAAGTGGCGCAGCTGTGC	1506
Db	840	IleGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgLysHisLeuCys	859
QY	1507	AAGCTGCTCGCGCGCGCAAGGCCCTTGACCGACATCGTCCCTCGTACCAGAGAGCGCGAG	1566
Db	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu	879
QY	1567	CTGGAGCTGGCGAGAACCGCGAGATCTTGGCGGAGCCCGTGCACCGCGCTGTACTAGCAC	1626
Db	880	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	899
QY	1627	CCAGCAAGGACCTGTGTGGCCGAGATCCAGAGAGCGGCGCACGACGAGTGGACCTACCAG	1686
Db	900	ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln	919
QY	1687	ATCTACGAGGACCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCACCGCC	1746
Db	920	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	939
QY	1747	CACACCAAGAGTGAAGCAGCTGACCGAGGCGCGTGCAGAAATCGCCATGCCATGGAGCATC	1806
Db	940	HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle	959
QY	1807	GTGATCTGGGCAAGACCCCGCAAGTTCCGCTTCCGCTGCATCCAGAAAGAGACCTGGGAGAC	1866
Db	960	ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr	979
QY	1867	TGGTGACCGCACTACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCC	1926
Db	980	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
QY	1927	CCCTGTGTGAAGCTGTGGTACCAAGTGGAGAGAGGCCCATCATCATCGCGCGCGAGCCTTC	1986
Db	1000	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe	1019
QY	1987	TACGTGGAGCGCGCGCAACCGCAGAGCAAGATCGCAAGCGCCGGCTACGTGACCGCAC	2046
Db	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp	1039
QY	2047	CGGGGCGCGCAGAAAGATCTGTGAGCTGACCGGAGACCCACCAAGAGACCGAGCTGCAG	2106

```

Db 1040 LysGlyArgGlnYsIleValSerLeuAsnGluThrThrAenGlnLysAlaGluLeuGln 1059
Qy 2107 GCATCCAGCTGGCCCTGCAGCAGCAGCGGACGAGGTGAACATCGTGACCGACGACG 2166
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079
Qy 2167 TAGCCCTGGCATCATCCAGGCCAGCCGACAGAGCGAGAGCGAGCGTGGTGAACGAG 2226
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099
Qy 2227 ATCATCGAGCGCTGATCAGAGAGAGAGTGTACCTGAGCTGGGTGCCGCCCAAG 2286
Db 1100 IleIleGlnLeuIleIleLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119
Qy 2287 GCATCCGCGCGCACGACGACGATCGACGAAGCTGTGAGCAAGGCGCATCCGCAAGTGTG 2346
Db 1120 GlyIleGlyAenGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139
Qy 2347 TTCTGTGACGCGATCGAT 2364
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 10
Q9WFA0 PRELIMINARY; PRT; 1427 AA.
AC Q9WFA0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-poi polypeptide.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana.";
RL J. Virol. 73:4427-4432(1999).
CC -I- SIMILARITY: Belongs to peptidase family A2.
DR EMBL; AF110978; A017162.1; -.
DR PIR; S49086; S49086.
DR HSP; Q70622; IHW.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0015074; P:DNA integration; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn_N.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR009007; Pept_AspAsp.
DR InterPro; IPR001969; Pept_AspAS.
DR InterPro; IPR010999; Retroviral_matrix.
DR InterPro; IPR000071; Retroviral_p17.
DR InterPro; IPR008916; Retrov_capsid_C.

```

```

DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010659; RVT_connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00655; Rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF00098; zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIVMATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
DR PROSITE; PS0158; ZF_CCHC; 2.
DR AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 160503 MW; 24C410797550C0F8 CRC64;

Alignment Scores:
Pred. No.: 7.47e-148 Length: 1427
Score: 3931.00 Matches: 727
Percent Similarity: 96.95% Conservative: 35
Best Local Similarity: 92.49% Mismatches: 22
Query Match: 85.96% Indels: 2
DB: 2 Gaps: 1

US-09-610-313B-31 (1-2463) x Q9WFA0 (1-1427)
Qy 13 ATGCGCGAGCCATGACGACGAGCGCACCGCCCAACATCTCTGATGACGCGCAGCACTTC 72
Db 360 LeuAlaGluAlaMetSerGlnAlaAsnAsnThrAsnIleMetIleGlnArgSerAsnPhe 379
Qy 73 AAGGCGCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGCGCCACATCGCCGC 132
Db 380 LysGlyProArgArgSerValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399
Qy 133 AACTGCGCGCGCCCGCGCAGAGAGGCTGTGAGTGGCGGCAAGGAGGCGCCACCATG 192
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyGlnGluGlyHisGlnMet 419
Qy 193 AAGGACTGCACCGAGCGCCAGCGCAACTTCTCCGCGAGGACCTTCCCGCCAGGC 252
Db 420 LysAspCysThrGluArgGlnAlaAspPheArgGluAsnLeuAlaPheProGlnGly 439
Qy 253 AAGGCGCGGAGTTCCTCCAGCGAGAGAACCGCGCCCAACAGCCCCACCGCGCGAGCTG 312
Db 440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerGlyGluLeu 459
Qy 313 CAGTGGCGGCGGACACACCCCGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGTGAAC 372
Db 460 GlnValTrpGlyAspAsnProArgSerGluThrGlyAlaLysGlyGlnGlyThrPheAsn 479
Qy 373 TCCCCCAGATCACCTGTGGCAGCGCCCTGTGAGCATCAAGGTGGCGCGCGCGCATC 432
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnVal 499
Qy 433 AAGGAGCGCCTGTCTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCC 492
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspThrValLeuGluGluIleAsnLeuPro 519
Qy 493 GCGAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGCATC 552
Db 520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyr 539

```

QY 553 GACCAGATCCTGATCGAGATCTGCGCAAGAAGGCCATCGGCACCGTGTGATCGGCCCC 612  
 Db |||||  
 QY 540 AspGlnIleThrIleGluLeuCyGlyLysLysAlaIleGlyThrValLeuValGlyPro 559  
 Db |||||  
 QY 613 ACCCCGCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGCTGCAACCTGCACTTC 672  
 Db |||||  
 QY 560 ThrProValAsnIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPhe 579  
 Db |||||  
 QY 673 CCATCAGCCCATCGAGACCGTGCCTGCGCGTGAAGCTGAAGCCCGGCATGACCGCCCAAG 732  
 Db |||||  
 QY 580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
 Db |||||  
 QY 733 GTGAACGCTGGCCCTGACCGAGGAGAAGATCAAGCGCCCTGACGCCATCTCGCAGGAG 792  
 Db |||||  
 QY 600 ValLysGlnTrpProLeuThrGluGluLysLysAlaLeuThrAlaIleCysGluGlu 619  
 Db |||||  
 QY 793 ATGGAGAAGGAGCGCAAGATCAACAGATCGGCCCGCAGAACCCCTACAAACCCCGCTG 852  
 Db |||||  
 QY 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
 Db |||||  
 QY 853 TTCGCCATCAAGAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTG 912  
 Db |||||  
 QY 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
 Db |||||  
 QY 913 AACAGCGCACCCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCCAGCCCGCGCCCTG 972  
 Db |||||  
 QY 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProLisProAlaGlyLeu 679  
 Db |||||  
 QY 973 AAGAAGAAGAAGCGTGCAGCTGTGGAGCTGGCGCAGCCCTTACAGCTGCCCTCG 1032  
 Db |||||  
 QY 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
 Db |||||  
 QY 1033 GACGAGGACTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCCCC 1092  
 Db |||||  
 QY 700 AspGluSerPheArgLysTrpAlaPheThrIleProSerIleAsnAsnAlaThrPro 719  
 Db |||||  
 QY 1093 GCATCCGCTACCAAGTACAGCTGCTGCCAGGCTGGAAGGGCAGCCCAAGCACTTC 1152  
 Db |||||  
 QY 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
 Db |||||  
 QY 1153 CAGAGCAGCATGACCAAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATC 1212  
 Db |||||  
 QY 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759  
 Db |||||  
 QY 1213 TACCAG-----GCCCCCTGACGTGGCAGGACCTGGAGATCGGCCAGCCGCC 1266  
 Db |||||  
 QY 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
 Db |||||  
 QY 1267 AAGATCGAGGAGTGGCGAAGCACCTGCTGCGTGGGGCTTCAACACCCCGCAGCAAGAG 1326  
 Db |||||  
 QY 780 LysIleGluGluLeuArgThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799  
 Db |||||  
 QY 1327 CACCAGAAGGAGCCCTCTCTGTGGATGGGTACGAGCTGCACCCCGCAGAGTGAC 1386  
 Db |||||  
 QY 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
 Db |||||  
 QY 1387 GTGCAGCCCATCGAGTGGCCGAGAGAGAGAGTGGACCGTGAACACATCCAGAGCTG 1446  
 Db |||||  
 QY 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
 Db |||||  
 QY 1447 GTGGGCAAGCTGAACCTGGCGCACCGATCTACCCCGGCATCAAGTGGCGCAGCTGTGC 1506  
 Db |||||  
 QY 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuCys 859  
 Db |||||  
 QY 1507 AAGCTGCTGCGCGCCCAAGCCCTGACCGCATCTGCCCCCTGACCGAGGAGCGCAG 1566  
 Db |||||  
 QY 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879  
 Db |||||  
 QY 1567 CTGAGCTGGCCGAGAACCGGAGATCTTCGCGGAGCCCGTGCACCGCTGACTACGAC 1626  
 Db |||||  
 QY 880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 899  
 Db |||||

QY 1627 CCCCAAGAGGAGCTGTGTGCTCGAGATCCAGAACAGGGCCACGACGAGTGGACCTACCA 1686  
 Db |||||  
 QY 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
 Db |||||  
 QY 1687 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAGATGGCACCAGCC 1746  
 Db |||||  
 QY 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
 Db |||||  
 QY 1747 CACACCAAGCGTGAAGCAGCTGACCGAGGCGCTGCAAGATCGCCATCGAGAGCATC 1806  
 Db |||||  
 QY 940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959  
 Db |||||  
 QY 1807 GTGATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGAGACCTGGGAGACC 1866  
 Db |||||  
 QY 960 ValLeuTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
 Db |||||  
 QY 1867 TGGTGACCGACTACTGGCAGGCCACTCGATCCCGAGTGGAGTTCGTGACACCCCC 1926  
 Db |||||  
 QY 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProLysTrpLysTrpPheValAsnThrPro 999  
 Db |||||  
 QY 1927 CCCCTGGTGAAGCTGTGTACCGAGTGGAGAGGAGCCCATCATCGGCCCGCAGACTTC 1986  
 Db |||||  
 QY 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe 1019  
 Db |||||  
 QY 1987 TACGTGGACGGCGCCCAACCCGAGACCAAGATCGCAAGCGCGCTTACGTGACCGAC 2046  
 Db |||||  
 QY 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyCysValThrAsp 1039  
 Db |||||  
 QY 2047 CGGGCGCGCAGAAGATCGTGAGCTGACCGAGACCCAGACCAACAGAGACCGAGTGCAG 2106  
 Db |||||  
 QY 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
 Db |||||  
 QY 2107 GCATCCAGCTGCGCTGCAGGACAGCGCAGAGTGAACATCGTGACCGACCGACAG 2166  
 Db |||||  
 QY 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079  
 Db |||||  
 QY 2167 TACGCCCTGGGCATCATCCAGCCCGCAGCCGCAAGAGCAGAGCGAGTGTGTGAACCA 2226  
 Db |||||  
 QY 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
 Db |||||  
 QY 2227 ATCATCGAGCAGCTGATCAAGAGGAGAGGTGTACCTGAGTGGTGGCCGCCCAAGAG 2286  
 Db |||||  
 QY 1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
 Db |||||  
 QY 2287 GCATCCGCGGCAACAGCAGCATCGACAGCTGTGTGAGCAAGCGCATCCGCAAGGTGCTG 2346  
 Db |||||  
 QY 1120 GlyIleGlyGlyHisGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
 Db |||||  
 QY 2347 TTCCTGGACGGCATCGAT 2364  
 Db |||||  
 QY 1140 PheLeuAspGlyIleAsp 1145  
 Db |||||

RESULT 11

Q9WF65 PRELIMINARY; PRT; 1427 AA.  
 ID Q9WF65;  
 AC Q9WF65;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Gag-pol polyprotein.  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99214383; PubMed=10196340;  
 RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
 RA Foley B.T., Ntung'u T.P., Rahama M., Makhema M.J., Marlink R.,  
 RA Essex M.;  
 RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
 RT virus type 1 subtype C: a set of 23 full-length clones from  
 RT Botswana.";  
 RL J. Virol. 73:4427-4432(1999).











Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499  
Qy 433 AAGGAGCCCTCTGGACACCGCGCGCCGACACACCGTGTGGAGGATGAGCTGCC 492  
Db 500 ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro 519  
Qy 493 GGCAGTGGAGCCCAAGATATCGCGGCGATCGCGCTTCATCAAGGTGCGCGATAC 552  
Db 520 GlyIleTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrp 539  
Qy 553 GACCAGATCTCTGATCGAGATCTGGCGCAAGAGCCATCGCGCACCGTCTGATCGGCC 612  
Db 540 GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro 559  
Qy 613 ACCCCGCTGAACATCATCGCCGCGCAACATCTGACCCAGCTGGCTGCACCTGAACATTC 672  
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe 579  
Qy 673 CCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCGCGCATGGACGCGCCCAAG 732  
Db 580 ProfileSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy 733 GTGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAsp 619  
Qy 793 ATGAGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAGAACCCCTACAAACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
Qy 853 TTCGCCATCAAGAGAGGACGAGCAGCACCGAGTGGCGGCAAGCTGTGGACTTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAAAGCGCACCCAGGACTCTGGAGGTGAGCTGGGCATCCCGCCCGCGGCGT 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAGAGAGCGTGCACCGTCTGGACGTGGCGCAGCGCTACTTTCAGCGTCCCGCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGACTTCCGCAAGTACACCGCTTACCATCCCGACGATCAACAGAGACCCCG 1092  
Db 700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719  
Qy 1093 GGCATCCGCTACCAAGTCAACGTCTGCCAGGCTGGAGGCGAGGCGCGCGATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCTCGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759  
Qy 1213 TACAG-----GCCCGCTGTAGTGGCGAGCAGCTGGAGATCGGCGACACCGCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCGAGGAGCTGCGCAGCACCTGCTGGCTGGGGCTTCACCGCCCGCAAGAG 1326  
Db 780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys 799  
Qy 1327 CACCAGAGAGGAGCCCGCTTCTCTGTGGATGGCTACGAGCTGCACCGCGCAAGTGGACC 1386  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1387 GTGAGCCCATCGAGCTGCCGAGAGAGAGCTGGACCGGTGAACGATCCAGAGCTG 1446  
Db 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1447 GTGGCGAGCTGAATCGGCGCAGCAGATCTACCGCGCATCAAGGTGCGCGCTGTC 1506  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValArgHisLeuCys 859

Qy 1507 AAGCTGTGCGCGCGCCCAAGCGCCCTGACCGACATCGTGCCTGTGACCGAGAGCGCCAG 1566  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
Qy 1567 CTGAGCTGCGCGAGAACCGCGAGATCTCTGCGCGAGCCGCTGACCGCGGTGTACTACGAC 1626  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 899  
Qy 1627 CCCAGCAAGACCTGTGTGGCGCGAGATCCAGAAAGCAGGCGCCACGACCTGACCTTACCAG 1686  
Db 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrGln 919  
Qy 1687 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGGCAAGATCGCACCCGCC 1746  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1747 CACACCAACGACGAGGAGCTGACCGAGCCCTGCGAGAGATCGCCATCGAGAGCATC 1806  
Db 940 HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle 959  
Qy 1807 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAAAGGAGACCTGGAGAGCC 1866  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979  
Qy 1867 TGTGGACCGACTACTGTGGCAGGCCCTGTGATCCCGAGTCCCGAGTTCGTGAACACCCGCC 1926  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1927 CCCCTGTGTGAAGCTGTGTACCGCTGAGAGAGAGCCCATCATCGCGCGCGAGACCTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
Qy 1987 TACGTGACGCGCGCGCAACCGGAGACCAAGATCGCAAGGCGGCTAGCTGACCCGAC 2046  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn 1039  
Qy 2047 CGGCGCGCGCAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2106  
Db 1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
Qy 2107 GCNATCCAGCTGCGCGCGCGAGGAGCGCGAGCGGCTGAACATCGTGACGAGCGCAG 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2167 TACGCGCTGGCATCATCCAGGCGCGCGCAGCAAGAGCGAGCGAGCTGGTGAACCCAG 2226  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
Qy 2227 ATCATCGAGCAGCTGTATCAAGAGGAGAGGTGTACCTGAGCTGGTGCCTGCCCGCCACAG 2286  
Db 1100 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy 2287 GGCATCGCGCGCAGCAGCAGATCGACAGCTGTGTGACCAAGGCGCATCCGCAAGTGTG 2346  
Db 1120 GlyIleGlyGlyAsnGluValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy 2347 TTCCTGGACGCGCATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 14  
Q8ADX5 PRELIMINARY; PRT; 1427 AA.  
ID Q8ADX5  
AC Q8ADX5  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol fusion polypeptide (fragment).  
GN Name=gag-pol;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;



QY 793 ATGGAGAGGAGGCAAGATCACCAGATCGGCGCCGAGACCCCTCAACACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639  
QY 853 TTCCGCATCAAGAAGAGAGCAGCACCAGTGGCGCAAGCTGGTGCATCTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysIleTyrArgLysLeuValAspPheArgGluLeu 659  
QY 913 AACAGCCACCCAGGACTTCTGGAGGTGAGCTGGGCATCCCCCAACCCGCGCGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
QY 973 AAGAGAGAGAGCGCTGACCGTCTGACCTGGCGGCGGCTACTCTCAGCGTGCCTG 1032  
Db 680 LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
QY 1033 GACGAGAGCTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAAAGAGACCC 1092  
Db 700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GGCATCCGCTACCAAGTACAAGCTGTGCTGCCAGGGCTGGAAGGCGAGCCCGAGCATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTyrLysGlySerProAlaIlePhe 739  
QY 1153 CAGAGCAGCATGACCAAGATCTGGAGCCCTTCCGCGCGCCGCAACCCCGAGATCTGATC 1212  
Db 740 GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759  
QY 1213 TACCAG-----GCCCGCTGTACTGGCGGACGACCTGGAGATCGGCGACCGCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779  
QY 1267 AAGATCGAGGAGCTGCGCAAGCAGCTGCTGGCTGGGCTTCCACACCCCGCAAGAG 1326  
Db 780 LysIleGluGluLeuArgGluHisLeuLeuArgTyrGlyPheThrProAspLysLys 799  
QY 1327 CACGAGAGGAGCCCGCTTCTGTGATGGCTACGAGCTGCACCCCGACAGTGAC 1386  
Db 800 HisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTyrThr 819  
QY 1387 GTGAGCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTG 1446  
Db 820 ValGlnProIleGlnLeuProGluLysAspSerThrValAsnAspIleGlnLysLeu 839  
QY 1447 GTGGGCAAGCTGAATCGGCGCAGCAGATCTACCCCGCATCAAGGTGCGCAGCTGTC 1506  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys 859  
QY 1507 AAGTGTGCGCGCGCGCAAGCCCTGACCGACATCTGCGCCCTGACCGAGGAGCCGAG 1566  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1567 CTGAGCTGGCGGAGAACCGGAGATCTGCGCGAGCCCGTGCACCGGCTGTACTACGAC 1626  
Db 880 LeuGluLeuAlaGluAsnArgGluLysLysGluProValHisGlyValTyrTyrAsp 899  
QY 1627 CCCAGCAAGCACTGTGGCGCAGATCCAGAGCAGGCGCCACGACCATGAGTGCACCTAC 1686  
Db 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpTyrGln 919  
QY 1687 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCGAGATCGCGACCGCC 1746  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939  
QY 1747 CACACCAAGCAGCTGAGCAGCTGACCGAGCGCTGCGAGATCGCGATCGCATGAGAGTATC 1806  
Db 940 HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959  
QY 1807 GTGATCTGGGCAAGACCCCAAGTCTCGCTGCGCCATCCAGAGAGAGACTGGGAGAC 1866  
Db 960 IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979  
QY 1867 TGGTGGACCGACTACTGCGCAGGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCC 1926

Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY 1927 CCCCTGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGCGCCGAGACCTTC 1986  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019  
QY 1987 TACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2046  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039  
QY 2047 CGGCGCGCGCAAGATCGTGCAGCTGACCGAGACCAACCAAGACCGAGCTGCGAG 2106  
Db 1040 ArgGlyArgGlnLysIleIleSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
QY 2107 GCCATCCAGCTGGCGCTGCAGCAGCGCGCAGCGAGTGAACATCGTGACCGACCGCCAG 2166  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValThrAspSerGln 1079  
QY 2167 TACGCTGGCGCATCATCCAGCGCGCGCGCGAGCGAGCGAGCGAGCTGTGAACCGAG 2226  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099  
QY 2227 ATCATCGAGCAGCTGATCAAGAGAGAGCTGTACCTGAGCTGGTGGTCCCGCCACAG 2286  
Db 1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2287 GGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGCGCGCATCCGCAAGTGTG 2346  
Db 1120 GlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
QY 2347 TTCTCGAGCGCATCGAT 2364  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 15  
Q9WF79 PRELIMINARY; PRT; 1425 AA.  
ID Q9WF79  
AC Q9WF79;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polypeptide.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana";  
RL J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL: AF110969; AAD17087.1; -;  
DR HSSP: P24736; INCP.  
DR GO: GO:0019012; C:virion; IEA.  
DR GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO: GO:0003677; F:DNA binding; IEA.  
DR GO: GO:0008907; F:integrase activity; IEA.  
DR GO: GO:0008233; F:peptidase activity; IEA.  
DR GO: GO:0004523; F:ribonuclease H activity; IEA.  
DR GO: GO:0003723; F:RNA binding; IEA.  
DR GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO: GO:0016740; F:transferase activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0015074; P:DNA integration; IEA.  
DR GO: GO:0006310; P:DNA recombination; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.



Db	838	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys	857
Qy	1507	AAGTGTGTCGGCGCCCAAGCCCTGACCATCGTGGCCCTGACCCAGGAGGCGCGAG	1566
Db	858	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProProThrGluGluAlaGlu	877
Qy	1567	CTGAGCTGGCGCGAGAACCGGAGATCTCTGGCGAGCCCGTGCACGGCGTGTACTACGAC	1626
Db	878	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAsp	897
Qy	1627	CCCAGCAGGACCTGGTGGCGAGATCCAGAGCAGGCGCCAGCAGCAGTGGACCTACCAG	1686
Db	898	ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspGlnTrpThrTyrGln	917
Qy	1687	ATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATCGCACCGCC	1746
Db	918	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla	937
Qy	1747	CACACCAACGAGTGAAGCAGCTGACCGAGCCGTCGACAGATCGCCATCGAGAGCATC	1806
Db	938	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle	957
Qy	1807	GTGATCTGGCGCAGACCCCAAGTTCGGCTGCCCTCCATCCAGAGGAGACCTGGGAGAC	1866
Db	958	ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla	977
Qy	1867	TGGTGGACCGACTACTGGCAGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCC	1926
Db	978	TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	997
Qy	1927	CCCTGTGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGGCGCCGAGACCTTC	1986
Db	998	ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe	1017
Qy	1987	TACGTGAGCGCGCGCCCAACCGGAGACCAAGATCCGCAAGCGCGCTAGTACCGGAC	2046
Db	1018	TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp	1037
Qy	2047	CGGCGCGGCAAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG	2106
Db	1038	LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln	1057
Qy	2107	GCCATCCAGCTGGCCCTGACGAGACAGCGGCGAGGTGAACATCGTGACCGAGCCAG	2166
Db	1058	AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln	1077
Qy	2167	TACGCCCTGGGCATCATCCAGGCCCGCCAGACAGAGCGAGGAGCTGGTGAACCCAG	2226
Db	1078	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln	1097
Qy	2227	ATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACTCTGAGTGGTGGTCCCGCCACAAG	2286
Db	1098	IleIleGlnLeuLeuIleLysLysGlnLysValTyrLeuSerTrpValProAlaHisLys	1117
Qy	2287	GGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAAGGCGCATCCGCAAGGTGCTG	2346
Db	1118	GlyIleGlyLysAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu	1137
Qy	2347	TTCCTGGACGCATCGAT	2364
Db	1138	PheLeuAspGlyIleAsp	1143

Search completed: June 2, 2005, 05:04:08  
Job time : 290.667 secs



***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 04:23:54 ; Search time 153.167 Seconds  
(without alignments)

11117.427 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 4573

Sequence: 1 gtcagccaccatggccga.....gggctagcaccgtgaattc 2463

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters: 2931222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09610313/runat\_31052005\_155138\_15193/app\_query.fasta\_1.7893  
-DB=Published Applications AA -QFMT=fastan -SURFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCU=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09610313 @CGN 1 1 550 @runat\_31052005\_155138\_15193  
-NCPU=6 -ICPU=3 -NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5  
-FGAPOP=6 -FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

\* Query  
Result

No.	Score	Match	Length	DB	ID	Description
1	3570	78.1	998	16	US-10-332-413-4	Sequence 4, Appli
2	3523.5	77.1	850	16	US-10-332-413-10	Sequence 10, Appl
3	3497	76.5	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3475	76.0	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3452	75.5	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3452	75.5	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3430	75.0	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3428	75.0	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3427	74.9	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3427	74.9	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3419	74.8	995	15	US-10-296-734-2	Sequence 2, Appli
12	3405	74.5	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3251.5	71.1	1006	15	US-10-296-734-1471	Sequence 1471, Ap
14	3183.5	69.6	1014	14	US-10-301-661A-6	Sequence 6, Appli
15	3119.5	68.2	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3119.5	68.2	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3119.5	68.2	1350	17	US-10-636-730-35	Sequence 35, Appl
18	2822.5	61.7	875	10	US-09-952-060-6	Sequence 6, Appli
19	2822.5	61.7	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2822.5	61.7	875	16	US-10-380-641-6	Sequence 6, Appli
21	2822.5	61.7	875	17	US-10-636-730-6	Sequence 6, Appli
22	2817	61.6	850	10	US-09-952-060-2	Sequence 2, Appli
23	2817	61.6	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2817	61.6	850	16	US-10-380-641-2	Sequence 2, Appli
25	2817	61.6	850	17	US-10-636-730-2	Sequence 2, Appli
26	2798.5	61.2	875	10	US-09-952-060-8	Sequence 8, Appli
27	2798.5	61.2	875	15	US-10-168-217A-8	Sequence 8, Appli
28	2798.5	61.2	875	16	US-10-380-641-8	Sequence 8, Appli
29	2798.5	61.2	875	17	US-10-636-730-8	Sequence 8, Appli
30	2793	61.1	850	10	US-09-952-060-4	Sequence 4, Appli
31	2793	61.1	850	15	US-10-168-217A-4	Sequence 4, Appli
32	2793	61.1	850	16	US-10-380-641-4	Sequence 4, Appli
33	2793	61.1	850	17	US-10-636-730-4	Sequence 4, Appli
34	2789	61.0	582	9	US-09-735-487-14	Sequence 14, Appl
35	2789	61.0	582	14	US-10-342-188-14	Sequence 14, Appl
36	2785	60.9	560	15	US-10-399-920-2	Sequence 2, Appli
37	2783	60.9	561	14	US-10-059-271-87	Sequence 87, Appl
38	2777	60.7	562	14	US-10-102-622-10	Sequence 10, Appl
39	2758	60.3	560	14	US-10-205-641-1	Sequence 1, Appli
40	2748.5	60.1	561	14	US-10-102-622-12	Sequence 12, Appl
41	2736	59.8	560	11	US-09-725-652-1	Sequence 1, Appli
42	2702	59.1	546	8	US-08-808-031A-29	Sequence 29, Appl
43	2689	58.8	979	14	US-10-271-181B-118	Sequence 118, Appl
44	2588	56.6	851	14	US-10-059-271-88	Sequence 88, Appl
45	2571	56.2	999	15	US-10-346-000A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-10-332-413-4  
; Sequence 4, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdoal  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; TITLE OF INVENTION: Subtype Isolates, Their Genes nd Modifications and Derivatives T  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924

; PRIOR FILING DATE: 2000-09-15 ; NUMBER OF SEQ ID NOS: 32 ; SOFTWARE: Patent In version 3.1 ; SEQ ID NO 4 ; LENGTH: 998 ; TYPE: PRT ; ORGANISM: Human immunodeficiency virus type 1 ; US-10-332-413-4									
Alignment Scores: Pred. No.: 1,566-165 Length: 998 Score: 3570.00 Matches: 666 Percent Similarity: 97.07% Conservative: 29 Best Local Similarity: 93.02% Mismatches: 19 Query Match: 78.07% Indels: 2 DB: 16 Gaps: 1									
US-09-610-313B-31 (1-2463) x US-10-332-413-4 (1-998)									
Qy	223	TTCCGCGAGACCTGGCGCTTCCGCCCAAGGCGCGAGTTCCTCCAGCGAGCAGAAC	282						
Db	1	PheArgGluAsnLeuAlaPheProGlnGlyGluAlaAargGluPheProSerGluGlnThr	20						
Qy	283	CGGCGCAAGCCCCACCGCGGAGCTGCGAGTGGCGGCGCACAACCCCGCAGCGAG	342						
Db	21	ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu	40						
Qy	343	CGCGCGCCGAGCGCGAGGSCACCCTGAACCTTCCCCAGATCACCTGTGGCAGCGCCCC	402						
Db	41	ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro	60						
Qy	403	CTGTGTAGCATCAAGGTGGCGCGCCAGATCAAGAGGCGCCTCTGGACACCGCGCGAC	462						
Db	61	LeuValSerIleLysIleGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp	80						
Qy	463	GACACCGTGTGAGAGATGAGCCTGCGCGGCAAGTGAACCCACAAGATCGCGCGC	522						
Db	81	AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly	100						
Qy	523	ATCGCGCGTTCATCAAGTGGCCAGTAGTACGACCATCTCATCGAGATCTGGCGGCAAG	582						
Db	101	IleGlyGlyPheIleLysValArgGlnThrAspGlnIleLeuIleGluIleCysGlyLys	120						
Qy	583	AAGGCCATCGGCACCGTGTGATCGCGCGCCACCCCGTGAACATCATCGCGCGCAACATG	642						
Db	121	LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet	140						
Qy	643	CTGACCCAGTGGGTGACCCCTGAACCTTCCCATCAGCCCCCATCGAGACCGTGGCCGTG	702						
Db	141	LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal	160						
Qy	703	ARGCTGAAGCCCGCATGGACGCGCCCAAGGTGAAGCTGCGCCTGACCGAGGAGAAG	762						
Db	161	LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlnGluLys	180						
Qy	763	ATCAAGCCCTGACCGCCATCTCGAGGAGATGGAAAGGCGGAAGATCACCAAGATC	822						
Db	181	IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle	200						
Qy	823	GGCCCCGAGAACCCCTACAACAACCCCGTGTTCGCCCATCAAGAAGAGGACAGCAACAG	882						
Db	201	GlyProGluAsnProThrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys	220						
Qy	883	TGGCGCAAGCTGTGGACTTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGGGAGGTG	942						
Db	221	TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal	240						
Qy	943	CAGCTGGGCATCCCCACCCCGCGCTGAAGAAGAAGAAGAGGCTGACCGTGTGGAC	1002						
Db	241	GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp	260						
Qy	1003	GTGGGCGACGCGCTACTTCAGCGTGCCTCGGACGAGGACTTCGCGAAGTACACCGCCTTC	1062						

QY 2137 AGCAGGTGAACATCTGTGACCGACGACGAGTACGCCCTCGGCGATCATCCAGGCGCCAGCCC 2196  
Db 641 SerGluValAsnIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660  
QY 2197 GACAAGAGCAGAGCGAGCTGGTGAACACAGATCATCCAGCAGCTGATCAAGAGGAGAG 2256  
Db 661 AspArgSerGluSerGluLeuValAsnGlnIleIleGlnLeuIleIleIleIleIleIleIle 680  
QY 2257 GTGTACCTGAGCTGGTGTGCGCCGCCCAACAGGCGATCGCGCGCAACAGCAGATCGACAAG 2316  
Db 681 ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700  
QY 2317 CTGGTGAGCAAGGCGATCCCGAAGGTGCTGCTTCTGAGCGGCGATCGAT 2364  
Db 701 LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716

RESULT 2  
US-10-332-413-10  
; Sequence 10, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swanson, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williams, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Selected H  
; TITLE OF INVENTION: Subtype Isolates, Their Genes and Modifications and Derivatives  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 850  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-10-332-413-10

Alignment Scores:  
Pred. No.: 2.8e-163 Length: 850  
Score: 3523.50 Matches: 668  
Percent Similarity: 91.05% Conservative: 34  
Best Local Similarity: 86.64% Mismatches: 44  
Query Match: 77.05% Indels: 25  
DB: 16 Gaps: 5

US-09-610-313b-31 (1-2463) x US-10-332-413-10 (1-850)

QY 97 TGCTTCACTCGCGAGGAGGCGCCACATCGCCGCGCACTGCGCGCCCGCCGCAAGAG 156  
Db 84 CysAlaThrValGlyLys-----GlyAspArgCysGlyProLeuArgTyrTyr 99  
QY 157 GGCTGCTGGAAG-----TGCGCAAGGAGGCGCCACCATG 192  
Db 100 AlaSerTrpArgLysGlyAspValLeuGlnGlyAspValGlyArgGlnGlyPheProSer 119  
QY 193 AAGGATGACCGAGGCGCCAG-----GCCAATCTTCTCCGCGAGGACCTG--- 237  
Db 120 HisAspValValLysArgArgProValProSerLeuHisAlaCysArgSerThrLeuGlu 139  
QY 238 -----GCCTTCCCGAGGCGCCAGCGCGCGAGTCCCGCGAGGAG 279  
Db 140 AspProArgValProSerPheProGlnGlyProAlaArgGlnPheProSerGluGln 159

QY 280 AACCGCGCCAAACAGCCGCCACAGCGCGAGCTGCGCGCGCGCAACACCCCGCGAGC 339  
Db 160 ThrArgAlaAsnSerProThrSerArgGluLeuGlnValArgAspAsnProArgSer 179  
QY 340 GAGCGCGCGCGCGAGCGCGCCAGCCCTCAATCTCCCGCCAGATCACCCTGTGCGAGCGC 399  
Db 180 GluThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArg 199  
QY 400 CCCTCTGTGAGCATCAAGGTGGCGCGCCAGATCAAGAGGAGCGCCCTGTGCGAGCACCGCGGCC 459  
Db 200 ProLeuValSerIleLysIleGlyGlnThrArgGluAlaLeuLeuAspThrGlyAla 219  
QY 460 GAGCAGACCGTGTGAGGAGATGAGCTTCCCGCGCAAGTGGAGGCCCAAGATGATCGGC 519  
Db 220 AspAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 239  
QY 520 GGCATCGCGCGCTCATCAAGGTGGCGCGCCAGTACGACGACGATCCTGATCGAGATCGCGC 579  
Db 240 GlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGly 259  
QY 580 AAGAAGCGCCATCGCGCCCGCTGTGATCGCGCCCGCCAGCGCGCGCGCGCGCGCGCGCGCG 639  
Db 260 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 279  
QY 640 ATGCTGACCGAGTGGGCTGCACCCCTGAATCTTCCCATCAGCCCCATCGAGACCGTGC 699  
Db 280 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 299  
QY 700 GTCAAGCTGAGCG 759  
Db 300 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluVal 319  
QY 760 AAGATCAAGCGCCCTGACCGCGCATCTGCGAGGAGATGGAGAGGAGGCGCGCGCGCGCGCG 819  
Db 320 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 339  
QY 820 ATCGCGCGCGAGAACCCCTACAACACCCCGCTGTGCGCATCAAGAGAGAGGAGGAGCAC 879  
Db 340 IleGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysGluAspSerThr 359  
QY 880 AAGTGGCGCGAGTGGTGGACTTCCGCGAGCTGAACAGCGCGCGCGCGCGCGCGCGCGCG 939  
Db 360 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 379  
QY 940 GTGAGCTGGGCGATCCCG 999  
Db 380 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu 399  
QY 1000 GAGTGGCGCGCGCGCTTACTTCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059  
Db 400 AspValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAla 419  
QY 1060 TTCCATCCCGAGCATCAACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119  
Db 420 PheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeu 439  
QY 1120 CCGCGCGCGTGGAGGCG 1179  
Db 440 ProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSerMetThrLysIleLeuGlu 459  
QY 1180 CCCTTCCCG 1233  
Db 460 ProPheArgAlaLysAsnProGluIleValIleTyrGlnTyrMetAlaLeuTyrVal 479  
QY 1234 GGCGAGCGCTGAGATCG 1293  
Db 480 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHisLeu 499  
QY 1294 CTGCGCTGGGCTTCAACCG 1353  
Db 500 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeuTrp 519

```
QY 1354 ATGGCGTACGAGTCACCCCAAGTGGACCGTCAGCCCATCGAGCTGCCCGAGAAG 1413
Db |||||TyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLys 539
QY 1414 GAGAGCTGGACCGTGAACAGCATCCAGAAGCTGGTGGCAAGCTGAACCTGGCGCAGCCAG 1473
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 559
QY 1474 ATCTACCCCGCATCAAGTGGCCGACCTGTGCAAGCTGTGCGCGCGCCAGAGCCCTG 1533
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 579
QY 1534 ACCGACATCTGCCCTTCAGCCGAGGCGCGAGCTGGAGCTGGCCGAGAACCCGAGATC 1593
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 599
QY 1594 CTGCGGAGCCCGTGCACCGCGTGTACTACGACCCCAAGGACCTGGTGGCCGAGATC 1653
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 619
QY 1654 CAGAAGCAGGCGCACGACAGTGGACCTACAGATCTACAGAGCGCTTCAAGACCTG 1713
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 639
QY 1714 AAGACCGGCAAGTACGCCAAGATCGCCACCGCCACACACGACGTAAGCAGCTGACC 1773
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 659
QY 1774 GAGCGCGTGCAGAGATCGCCATCGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTC 1833
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 679
QY 1834 CGCTGCCCATTCAGAAGGAGACTGGGAGACCTGTGGACCACTACTGGCAGGCGCAC 1893
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 699
QY 1894 TGGATCCCGAGTGGAGTTCGTGAACACCCCGCTGGTGAAGCTGGTACCAAGCTG 1953
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 719
QY 1954 GAGAAGGAGCCCATCATCGCGCGCAGACCTTCTACGTGGACCGCGCGCCCAACCCGAG 2013
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 739
QY 2014 ACCAAGTCCGCAAGCGCGCTACGTACCGACCGACCGCGCGCGCGCAGAGATCGTAC 2073
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 759
QY 2074 ACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCGCTGCAGGACAGC 2133
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 779
QY 2134 GGCAGCGAGTGAACATCTGTACCGACGACGAGTACGCGCTGGCGCATCATCCAGGCCAG 2193
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 799
QY 2194 CCGCACAAGCAGCAGCAGCTGTGTAACACAGATCATCAGCAGCTGATCAAGAGAGAG 2253
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 819
QY 2254 AAGGTGTACCTGAGCTGGTGGCGCCGACCAAGGGCATCGCGCGCACGAGCAGATCGAC 2313
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 839
QY 2314 AAGCTGTGTAGCAGGCGCATCCGCAAGGTGCTG 2346
Db |||||TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeu 850
```

## RESULT 3

US-10-283-847-17

; Sequence 17, Application US/10283847

; Publication No. US20030162720A1

; GENERAL INFORMATION:

```
; APPLICANT: Nicolette, Charles A.
; APPLICANT: Walker, Bruce
; TITLE OF INVENTION: THERAPEUTIC ANTI-HIV (IV9) COMPOUNDS
; FILE REFERENCE: GZ 2112.00
; CURRENT APPLICATION NUMBER: US/10/283,847
; CURRENT FILING DATE: 2002-10-29
; PRIOR APPLICATION NUMBER: 60/345,116
; PRIOR FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus
; US-10-283-847-17
```

## Alignment Scores:

Pred. No.:	5,48e-162	Length:	1003
Score:	3497.00	Matches:	655
Percent Similarity:	95.56%	Conservative:	34
Best Local Similarity:	90.85%	Mismatches:	26
Query Match:	76.47%	Indels:	6
DB:	14	Gaps:	3

US-09-610-313B-31 (1-2463) x US-10-283-847-17 (1-1003)

```
QY 220 TTCTTCCGCGAGGACCTGGCTTCCCCAGGGCAAGGCCGCGAGTTCCTCCAGCGAGCAG 279
Db |||||PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCAAACAGCCCAACCCAGCCGCGAGCTGTCAGGTG-----CGCGCGCGCAACCC 333
Db |||||ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSer 40
QY 334 CGCAGCGAGCGCGCGCCGCGAGCGCCAGGCGCACCCCTG-----AAGTTCCTCCAGATCAC 387
Db |||||ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
QY 388 CTGTGGCAGCGCGCGCTGGTGGAGTCAAGTGGCGGCGCAGATCAGGAGGCGCTGCTG 447
Db |||||LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCCACACACCGTGTGTGGAGGAGATGAGCTGCGCGCGCAAGTGAAGCC 507
Db |||||AspThrGlyAlaAspAspThrValLeuGluMetSerLeuProGlyArgTrpLysPro 100
QY 508 AGATGATCGCGCGCATCGCGCGCTTTCATCAGTGGCGCGCAGTACGACGATCTGTGATC 567
Db |||||LysMetIleGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleLeuLeu 120
QY 568 GAGATCTGGCGCAAGAGGCGCATCGGCACCGCTGTATCGGCCCGCCACCGCGTGAACATC 627
Db |||||GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGTGTGACCCAGCTGGGTGTGACCCCTGAACCTTCCCATCAGCCCCATC 687
Db |||||IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGTAAGTGAAGCCCGCATGGACGCGCGCATGGAGCGGCCCAAGGTGAAGCAGTG 747
Db |||||GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGNAGATCAAGGCGCTTCAGCCCATCTGCGAGGAGATGGAGAGGAGGGGC 807
Db |||||LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCCCGCGAGAACCCCTACACACCCCGCTGTTCGCATCAAGAAG 867
Db |||||LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGTGGATCTCCGGGAGCTGAACAGCCGACCCAG 927
Db |||||LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
```

QY 928 GACTTCTGGAGGTGCTAGCTGGGATCCCCCAACCCCGCGCTCTGAAGAAAGAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyLeuProHisProAlaGlyLeuLeuLysLysSer 260  
QY 988 GTGACCGTGTGGAGCGGCGAGCCCTACTTTCAGCGTCCCTGGAGGAGCTTCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspGluAspPheArg 280  
QY 1048 AAGTACACCGCTTCCACCATCCAGCATCAACACGAGACCCCGGATCCGCTACCCAG 1107  
Db 281 LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 300  
QY 1108 TACAACGTGCTGCCCGAGGCTGAAGGGCAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTCTGGAGCCCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACCAG- ----GCC 1221  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTrpGlnTrpMetAsp 340  
QY 1222 CCCCTGTACGTGGCAGCAGCTCGAGATCGCCGAGCAGCCCGCGCAAGATCGAGGAGCTG 1281  
Db 341 AspleuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu 360  
QY 1282 CGCAAGCACCTGTGCTGGGCTTCCACCAACCCCGCAAGAGCAGCAGAGAGGCC 1341  
Db 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrProAspLysLysHisGlnLysGluPro 380  
QY 1342 CCCTTCTGTGGATGGGCTAGAGCTGCACCCGCAAGTGGAGCGGCGACCCATCGAG 1401  
Db 381 ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
QY 1402 CTGCCCGAGAGGAGCTGGACCGTGAACACATCCAGAGCTGGTGGCAAGCTGAAC 1461  
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
QY 1462 TGGCCCGACCGAGATCTACCCCGCATCAAGTGGCGCAGCTGTCAAGCTGCTCGCGGC 1521  
Db 421 TrpAlaSerGlnIleTrpProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
QY 1522 GCCAAGCCCTGACCGCATCGTCCCTGACCGAGGCGCGAGCTGGAGCTGGCGGAG 1581  
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
QY 1582 AACCGCAGATCTCGCGAGCCCGTGCACCGCTGTACTACGACCCCGCAGCAAGCACTG 1641  
Db 461 AsnArgGluLeuLysGluProValHisGlyValTrpAspProSerLysAspLeu 480  
QY 1642 GTGCCCGAGATCCAGAGCGGCGCAGCACGAGTGGACCTTACCGATCTTACCGAGGCC 1701  
Db 481 IleAlaGluIleGlnLysGlnGlyGlnTrpThrTrpGlnIleTrpGlnGluPro 500  
QY 1702 TTCAGAACCTTGAGACCGCAAGTACGCCAAGATGGCCAGCGCCGCAACCAAGCACTG 1761  
Db 501 PheLysAsnLeuLysThrGlyLysTrpAlaArgMetArgGlyAlaHisThrAsnAspVal 520  
QY 1762 AAGCAGCTGACCGAGGCGGTGACAGATCCCATCGAGAGCATCTGTGATCTGGGGCAAG 1821  
Db 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTrpGlyLys 540  
QY 1822 ACCCCCAAGTTCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTTAC 1881  
Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTrp 560  
QY 1882 TGGCAGCGCAGCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941  
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
QY 1942 TGGTACCAGCTGGAGAGGAGCCCATCATCGGCCCGAGACCTTCTACGTGGAGCGGCC 2001  
Db 581 TrpTrpGlnLeuGluLysGluProIleValGlyAlaGluThrPheTrpValAspGlyAla 600

QY 2002 GCCAACCCGAGAGACCAAGATCGCAAGCCCGCTTACGTGACCGACCGGCGCGGAGAG 2061  
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTrpValThrAsnArgGlyArgGlnLys 620  
QY 2062 ATCTGTGACCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121  
Db 621 ValValThrLeuThrAspThrAsnGlnLysThrGluLeuGlnAlaIleTrpLeuAla 640  
QY 2122 CTGACGACGACCGCGAGCGAGTGAACATCGTGACCGACAGCCAGTACGCTTGGGCATC 2181  
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTrpAlaLeuGlyIle 660  
QY 2182 ATCCAGCCCGACCGCGCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGCAGTG 2241  
Db 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu 680  
QY 2242 ATCAAGAGGAGAGGTGTACTGAGCTGGGTGGCGCCGACCAAGGGCATCGCGGCAAC 2301  
Db 681 IleLysLysGluLysValTrpLeuAlaTrpValProAlaHisLysGlyIleGlyGln 700  
QY 2302 GAGCAGATCGCAAGCTGGTGAGCAAGGGCATCCGCAAGGTGTCTTCTTGGACGGCATC 2361  
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720  
QY 2362 GAT 2364  
Db 721 Asp 721  
RESULT 4  
US-10-634-165-9  
; Sequence 9, Application US/10634165  
; Publication No. US20050095581A1  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Nancy T  
; APPLICANT: Gallo, Robert C.  
; APPLICANT: Wong-Staal, Flossie  
; TITLE OF INVENTION: DETECTION OF HIV-1 DNA  
; FILE REFERENCE: 223695  
; CURRENT APPLICATION NUMBER: US/10/634,165  
; PRIOR FILING DATE: 2003-08-05  
; PRIOR APPLICATION NUMBER: 08/463,028  
; PRIOR FILING DATE: 1995-06-05  
; PRIOR APPLICATION NUMBER: 08/693,866  
; PRIOR FILING DATE: 1985-01-23  
; PRIOR APPLICATION NUMBER: 06/659,339  
; PRIOR FILING DATE: 1984-10-10  
; PRIOR APPLICATION NUMBER: 06/643,306  
; PRIOR FILING DATE: 1984-08-22  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 9  
; LENGTH: 1015  
; TYPE: PRT  
; ORGANISM: T cell leukemia-lymphoma virus (HTLV)  
; FEATURE:  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)..(1015)  
; OTHER INFORMATION: /note= "pol protein of HTLV-III"  
US-10-634-165-9  
Alignment Scores:  
Pred. No.: 6,42e-161 Length: 1015  
Score: 3475.00 Matches: 654  
Percent Similarity: 93.72% Conservative: 33  
Best Local Similarity: 89.22% Mismatches: 28  
Query Match: 75.99% Indels: 18  
DB: 17 Gaps: 4  
US-09-610-313B-31 (1-2463) x US-10-634-165-9 (1-1015)  
QY 220 TTCTTCCGAGGACCTGGCTTCCCGAGGGGAGGCGCGGAGTTTC----- 267  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20

```
QY 268 -----CCGAGGAGCAAGAACCGCGCCCAACAGCCCCCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGGCGCAACACCCCGCAGCGAGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGGACCCCTG-----AACTTCCCCCAGATCACCTGTGGCAGCGCCCCCTGGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGGTGGCGGCCAGATCAAGGAGGCCCTCTGGACACCGCGCGCCGACGACACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGAGGAGATGACCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY 532 TTCAATCAAGGTGGCGGACGATACGACCAAGATCCTGATCGAGATCTGCGGCAAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTGTGATCGGCCCCACCCCGTGAACATCATCGGCCGCAACATGCTGACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCAACCTGAACCTTCCCATCAGCCCATCGAGACCGTGGCCCGTGAAGCTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY 712 CCCGGCATGACGCCGCCAAGGTGAAGCAGTGGCCCTGACCCAGAGAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTACCGCCCATCTCGGAGGAGATGGAGAGGAGGCAAGATCACCAAGATCGGCCCGCAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluLysIleSerLysIleGlyProGlu 220
QY 832 AACCCCTTACAACACCCCGGTGTTCCCATCAAGAAGAAGACAGCACCAAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGGTGAATCTCGCGAGCTGAACACAGCGCACCCAGACTTCTGGAGGTGACGTGGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCGCCACCCCGCGCTGAAGAAGAAGAAGAGCGTGACCGTGTGGACGTGGCGAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTTCAGCGTGGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACACGAGACCCCGCGATCCGCTACAGTACAACTGCTGCGCCCGCAGGCGTG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGGGAGCCCGAGCATCTTCCAGAGCAGCATCACCAAGATCTGGAGCGCTTCCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPhePheLysLys 340
QY 1192 CGCAACCCCGAGATCGTGATCTACCAAG-----GCCCGCTGTACGTGGCGACGACCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
QY 1246 GAGATGGCGGCAGCACCGCGCCAGATCGAGGAGCTGGCAAGCACCTGTGCGCTGGCGC 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGly 380
```

RESULT 5

```
QY 1306 TTCACCAACCCCGCAAGAAGACCAAGAGAGAGCCCCCTTCTGTGGATGGGCTACGAG 1365
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1366 CTGCACCCCGCAACAAGTGGACCGTGTGAGCCCATCGAGCTGCCGAGAGAGAGAGCTGACC 1425
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1426 GTCAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCGCAGCCAGATCTACCCCGGC 1485
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
QY 1486 ATCAAGGTGGCGCGAGCTGTGCAAGTGTCTCGCGCGGCCCAAGCCCTGACGACATCGTG 1545
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1546 CCCTGACCGAGGAGCGAGCTGTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCGAGCCC 1605
Db 461 ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1606 GTGCACCGGTGTACTACGACCCCGCAGGACCTGTGGCCGAGATCTGCGCGAGATCCAGAGCAGGC 1665
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1666 CACGACCACTGGACCTTACCAGATCTACACGAGGACCTTCAAGAACCTGAAGACCGGCAAG 1725
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1726 TAGCCCAAGATGGGACCCGCCACCAACGACGAGTGAAGAGTGAACCGCGCGCTGCAG 1785
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1786 AAGATCGCCATCGAGAGCATCTGTGCGGCAAGACCCCAAGTTCGCGCTGCCCATC 1845
Db 541 LysIleThrThrGlnSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1846 CAGAAAGAGACCTGGGAGACCTGTGTGGACCGCACTACTGCGAGCGCACCTGGATCCCGAG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1906 TGGGATTCGTGAACACCCCGCTGGTGAAGCTGTGTGTACAGCTGGAGAGAGAGCCC 1965
Db 581 TrpGluPheValAsnThrProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1966 ATCATCGCGCGCAGACCTTCTACGTGGACGCGCGCCCAACCGCAGACCAAGATCGGC 2025
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCGCGCTTACGTGACCCAGCCGCGCGCGCAGAAATCGTGAAGCTGACCGAGACACC 2085
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2086 AACCAGAACCGAGCTGCAGGCCATCTCAGCTGGCCCTGCGAGACAGCGGACGAGGTG 2145
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATCTGCACCGACAGCAGTACCGCTGCGCATCATCCAGCCCGCAGCCCGCAGAGAGC 2205
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2206 GAGAGCGAGCTGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTG 2265
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2266 AGCTGGTGGCCCGCCCAAGGCGCATCGCGCGCAACAGCAGATCCAGAGCTGTGTGAGC 2325
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2326 AAGGGCATCCGCAAGCTGTCTTCTGGACGGCATCGAT 2364
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```





```
Db 481 ValHisGlyValTyrTyAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1666 CACGACAGTGGACCTACAGACTTACCAGAGCCCTTCAAGAACCTGAAGACCGGCAAG 1725
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1726 TAGCCAAAGATGGCAGCCGCCACCAACACGAGTGAAGCAGCTGACCGAGGCCGTGCAG 1785
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
QY 1786 AAGATCGCCATGGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATC 1845
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1846 CAGAAGAGACCTGGGAGACTGTGTGACCGACTACTGGCAGCCACCTGGATCCCGAG 1905
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1906 TGGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGTACAGCTGGAGAGAGGCC 1965
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTyrTrpGlnLeuGluLysGluPro 600
QY 1966 ATCATCGGCCCGAGACCTTCTACGTGACCGCGCCCAACCCCGAGACCAAGATCGGC 2025
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2026 AAGCCCGCTACGTGACCGACCGCGCGCCGAGAGATCGTGTGACCTGACCGAGCACCC 2085
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2086 AACGAGAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGCAGCGAGTG 2145
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2146 AACATCGTGACCGACGACCAAGTAGTACGCCCTGGGATCATCCAGGCCCGACCAAGAGC 2205
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2206 GAGAGCGAGCTGTGAACAGATCATCGACACCTGATCAAGAGGAGAGGTGTACCTG 2265
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2266 AGCTGGGTGCGCCGCCAACAGGCGCATCGCGGCAACGAGCAGATCGACAGCTGGTGAGC 2325
Db 701 AlaTrpValProIleHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720
QY 2326 AAGGCGATCCGCAAGGTGTCTTCTGTGACGGCATCGAT 2364
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

RESULT 6

```
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE OF INVENTION: AN IMMUNE RESPONSE
; FILE REFERENCE: 12804-005002
; CURRENT APPLICATION NUMBER: US/10/093, 953A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
```

```
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by construct of vaccine vector
; OTHER INFORMATION: pGAL and vaccine insert expressing clade B HIV-1
; OTHER INFORMATION: Gag-Pol
; US-10-093-953A-31

Alignment Scores:
Pred. No.:      8,24e-160      Length:      739
Score:          3452.00      Matches:      651
Percent Similarity: 93.45%      Conservative: 34
Best Local Similarity: 88.81%      Mismatches: 30
Query Match:      75.49%      Indels:      18
DB:              16          Gaps:       4

US-09-610-313B-31 (1-2463) x US-10-093-953A-31 (1-739)
QY 220 TTCTTCCGCGAGACTGTGCTTCCCTCCAGGGCAAGCCCGCGAGTTCCTCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGC-----GCCAACGCCCCACGAC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGAGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGGCACCGCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCGCGCGCGACCGTG 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGGTGGCGCGCGCAGATCAAGGAGCGCTCTGGACACCGCGCGCGCGCGACCGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly 120
QY 532 TTCATCAAGGTGGCGCGCGCAGTACACAGATCTGTGATCGAGATCTGCGCGCAAGAGGCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTCTGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACCTTCCCATCAGCCCCATCAGACCGCGCGCGCGCGCGCGCGCG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY 712 CCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCCATCTGCGAGGAGATGGAGAGGCGGCGCAAGATCACCAAGATCGCGCGCG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
```



QY 334 CGCAGCGAGCGCGCGAGCGCCGAGCGCACCTG-----AACTTCCCGCCAGATCAC 387  
 Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnProGlnIleThr 60  
 QY 388 CTGTGGCAGCGCCCGCTGTGTGAGCATCAAGGTGGCGCGCAGATCAAGAGAGCGCTGCTG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlyLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCCGACGACACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGAAGCC 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGCAGTACGACACGATCCTGATC 567  
 Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProIle 120  
 QY 568 GAGATCTGCGCGCAAGAGCGCATCGGCACCGTGTGATCGCGCCCGCCCGCGTGAACATC 627  
 Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGCGCGCAATGTGTACCCAGCTGGTGTGACCTGACCTGAATCTCCCATCAGCCCATC 687  
 Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnProIleSerProIle 160  
 QY 688 GAGACCGTGCCTGAGCTGAAGCCCGCGCATGAGCGCGCCCAAGGTGAGAGTGGCC 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGAGAGAGATCAAGCGCTGACCGCCATCTGCGAGGAGATGAGAGAGAGGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 ARGATCACCAAGATCGCGCCGAGAACCCCTACAACACCCCGTGTTCGCATCAAGAAG 867  
 Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGCAGACGACCAAGTGGCGCAGCTGGTGGATCTCCGCGAGCTGACAGCGCACCCAG 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240  
 QY 928 GACTTCTGGAGGTGACGTGGGCGATCCCCACCCCGCGCTGAGAGAGAGAGAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGACGTGGCGCAGCCTACTTACGCGTGCCTTGGACGAGGACTTCCGC 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
 QY 1048 AAGTACACCGCTTCACCATCCCGCAGCATCAACAGAGACCCCGCGCATCCGCTACCG 1107  
 Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln 300  
 QY 1108 TACAACTGTGCCCCAGGCGCTGAGGCGCAGCCAGCATCTTCCAGAGCAGCATGACC 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCCTGGAGCCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACACG-----GCC 1221  
 Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleTyrGlnTyrMetAsp 340  
 QY 1222 CCCTGTACCTGGCAGCGCTGAGATCGCGCAGCAGCCCGCGCAGATCGAGGAGCTG 1281  
 Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGCACCTGTGCTGGCGCTTCCACACCCCGCAGCAAGAGACACCAAGAGAGGAGCC 1341  
 Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCTTCTCTGTGGATGGCTACGAGCTGCACCCGACCAAGTGGACCGTGGACGCCATCGAG 1401  
 Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400

QY 1402 CTGCCCGCAGAGGAGAGCTGGACCGGTGAACACATCCAGAGCTGGTGGCGAAGCTCAAC 1461  
 Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 QY 1462 TGGCGCAGCCAGATCTACCCCGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGGC 1521  
 Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440  
 QY 1522 GCCAAGCCCTGACCGACATCGTGGCCCTGAGCCGAGAGGCGCGAGCTGGAGCTGGCCGAG 1581  
 Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460  
 QY 1582 AACCGCGAGATCTCGCGCAGCGCGCTGTACTTACACCCCGCAGCAAGAGCCTG 1641  
 Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480  
 QY 1642 GTGGCGCAGATCCAGAAAGCGGCCACGACCTAGTGTGACCTTACCCAGATCTACAGAGCC 1701  
 Db 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500  
 QY 1702 TTCAAGACCTGAAAGCCCGCAGTACGCGCAAGATCGGCACCGCCACACCAAGCGAGCTG 1761  
 Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgSerAlaHisThrAsnAspVal 520  
 QY 1762 AAGCAGCTGACCGAGCGCGTGCAGAAAGATCGCCATGGAGAGCATCGTGATCTCGGGCAAG 1821  
 Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540  
 QY 1822 ACCCCCAAGTTCGCTGCTCCATCCAGAAAGGAGACCTGGGAGACCTGGTGAGCCGACTAC 1881  
 Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpGluTyr 560  
 QY 1882 TGCAGCGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGTGAGAGCTG 1941  
 Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu 580  
 QY 1942 TGTGTACAGCTGAGAAAGAGCGCCATCATCGCGCGCGAGACCTTCTACGTGGACGCGGCC 2001  
 Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600  
 QY 2002 GCCAACCGCGAGACCAAGATCGGCAAGCGCGGTACTGTGACCCGACCGCGCGCGCGAGAG 2061  
 Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620  
 QY 2062 ATCGTGTGCTGACCGAGACCCACCAAGACCGAGCTGAGCGCGAGCCATCCAGCTGGCC 2121  
 Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640  
 QY 2122 CTCGAGCAGACCGCGCAGCGAGGTGAACATCTGACCGCAGCAGCCAGTACGCTGGGCATC 2181  
 Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
 QY 2182 ATCCAGCGCCCGCAGCCGACCAAGCGAGAGCGAGCTGGTGAACCCAGATCATCGAGCAGCTG 2241  
 Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
 QY 2242 ATCAAGAGAGAGAGTGTACTTACCTGAGCTGGTGGTCCCGCCCGCAGAGGCGCATCGCGGCAAC 2301  
 Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700  
 QY 2302 GAGCAGATCCAGCAAGCTGGTGAAGAGGCGCATCCGCAAGGTGTGTCTTCTGGACGCGCATC 2361  
 Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720  
 QY 2362 GAT 2364  
 Db 721 Glu 721

RESULT 8  
 US-10-325-468-35  
 ; Sequence 35, Application US/10325468  
 ; Publication No. US20040101823A1  
 ; GENERAL INFORMATION:

; APPLICANT: Soong, Nay Wei  
 ; APPLICANT: Pekrun, Katja  
 ; APPLICANT: Shibata, Riri  
 ; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
 ; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS  
 ; FILE REFERENCE: 0166.210US  
 ; CURRENT APPLICATION NUMBER: US/10/325,468  
 ; CURRENT FILING DATE: 2002-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/343,524  
 ; PRIOR FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 56  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 35  
 ; LENGTH: 1003  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol  
 ; US-10-325-468-35

Alignment Scores:  
 Pred. No.: 1,23e-158 Length: 1003  
 Score: 3428.00 Matches: 639  
 Percent Similarity: 94.73% Conservative: 44  
 Best Local Similarity: 88.63% Mismatches: 32  
 Query Match: 74.96% Indels: 6  
 DB: 16 Gaps: 4

US-09-610-313B-31 (1-2463) x US-10-325-468-35 (1-1003)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGGCAGGCGCGAGTTCCTCCAGCGAGCAG 279  
 DB 1 PhePheargGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20  
 QY 280 AACCGCGCCAAACGCCACCAGCGCGAGCTGCAGGTG---CGCGCGAC---AACCCC 333  
 DB 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgGlyAspAsnAsnSer 40  
 QY 334 CGCAGCAGGCGCGCGCGAGCGCCAGCGCACCTCGT-----AACTTCCCGCAGATCAC 387  
 DB 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
 QY 388 CTGTGGCAGGCGCCCTGCTGTGAGATCAAGTGGCGCGCGCAGATCAAGAGGCGCCTCTG 447  
 DB 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCGCAGCAGCACCGTCTCGAGGAGATGAGCTGCCCGCGCAAGTGGAAGCCC 507  
 DB 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgGtrPlyPro 100  
 QY 508 AAGATGATCGGCGGCATCGCGGCTTCATCAAGGTGGCGCAGTAGTACGACACGATCCTGATC 567  
 DB 101 LysMetIleGlyIleGlyGlyPheIleLysValargGlnTyArgGlnIleProIle 120  
 QY 568 GAGATCTGGCGCAAGAGGCATCGGCACCGTGTGATCGGCCCGCCCGCCCGCGTGAATC 627  
 DB 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGGCGCAACATCTGACCGACCTGCGGTGACCTGCACTTCCCATCAGCCCGCCATC 687  
 DB 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACCGTGGCGTGAAGCTGAAGCCCGCATGACCGCGCGCGCGCGCGCGCGCGCGCG 747  
 DB 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGAGAGAGATCAGGCGCCTGACCGCCCATCTGCGAGGAGATGAGAGAGGCGC 807  
 DB 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCACCAAGATCGGCGCGCGAGAACCTCTACACACCGCGCGCGCGCGCGCGCGCG 867  
 DB 201 LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 220

QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGAGCTTCCGCGAGCTGAAACAAGCGCACCCAG 927  
 DB 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240  
 QY 928 GACTTCTGGGAGGTGCAGCTGGGCATCCCGACCCCGCGCGCTCAAGAAGAAGAGAGC 987  
 DB 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGGACGTGGCGCAGCTTCTACGCGTCCCGCTGGAGAGAGACTTCGCG 1047  
 DB 261 ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg 280  
 QY 1048 AAGTACACCGCTTCACATCCCGCAGCATCAACACGAGACCCCGCGCATCCGCTACCCAG 1107  
 DB 281 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 300  
 QY 1108 TACAACCTGTGCTCCCGAGGCTGGAAGGCGAGCCCGACGATCTTCCAGAGCAGCATGACC 1167  
 DB 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCTGTGAGCGCTTCCGCGCGCCCAACCCCGAGATCGTGATCTACCGAG-----GCC 1221  
 DB 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyThrMetAsp 340  
 QY 1222 CCCCTGTACGTGGCGCAGCGACCTGGAGATCGCGCAGCACCGCGCCCAAGATCGAGAGCTG 1281  
 DB 341 AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGCACCTGTGCGCTGGCGCTTCCACCAACCCCGCAGAGACCAAGAGACCAAGAGAGCCC 1341  
 DB 361 ArgGlnHisLeuLeuTyTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCTGTGTGAGTGGCTACGAGCTGCACCCCGACCAAGTGGAGCGCTGCAGCCCATCCAG 1401  
 DB 381 ProPheLeuTrpMetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
 QY 1402 CTGCGCGCAAGAGAGAGCTGGACCGGTGAAGCAGCATCCAGAAAGCTGGTGGGCAAGCTGAAC 1461  
 DB 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 QY 1462 TGGCGCAGCAGATCTACCCCGCGCATCAAGTGGCGCGCAGCTGTGCAGAGCTCTCGCGGC 1521  
 DB 421 TrpAlaSerGlnIleTyAlaGlyIleLysValLysGlnLeuLysLysLeuLysGly 440  
 QY 1522 GCCAAGCCCTGACCGACATCGTCCCTCAGCAGGAGGCGCGAGCTGGAGCTGGCGCAG 1581  
 DB 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuLysAlaGlu 460  
 QY 1582 AACCGCAGATCTCGCGCAGCGCGCTGCACGGCGTGTACTACGACCCCGCAGCAAGAGCTG 1641  
 DB 461 AsnArgGluIleLeuLysGluProValHisGlyValTyArgProSerLysAspLeu 480  
 QY 1642 GTGGCGCAGATCCAGAGCAGGCGCCACGACAGTGGACCTACAGAGCTTACCAGAGCGCC 1701  
 DB 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyThrGlnIlePheGlnGluPro 500  
 QY 1702 TTCAAGAACCTGAAGACCGCGCAAGTACGCCAAGATGCCCGCGCGCGCGCGCGCGCGCG 1761  
 DB 501 PheLysAsnLeuLysThrGlyLysTyArgLysThrArgGlyAlaHisThrAsnAspVal 520  
 QY 1762 AAGCAGCTGACCGAGCGCGTGCAGAAAGATCCCATGAGAGCATCTGTGATCTGGGCGAAG 1821  
 DB 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540  
 QY 1822 ACCCCCAAGTTCGCTCCCATCCAGAGGAGCTGGGAGACCTGGTGACGCTGACGACCTAC 1881  
 DB 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTy 560  
 QY 1882 TGGCAGCGCACCTGGATCCCGCGAGTGGAGTTCGTGAACACCCCGCGCGCGCGCGCGCG 1941  
 DB 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580

```
QY 1942 TGGTACCAGCTGGAGAGAGCCCATCATCGGGCCGAGACCTTCTACGTGGACGGCGCC 2001
Db 581 TTPTyrglnleuGlulysGluProlleValGlyAlaGluThrPheTyValaspGlyAla 600
QY 2002 GCCAACCGCGAGACCAGATCGGCAAGGCGCGGTACGTGACCGACCGGGCGGGCAGAAG 2061
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyValThrSerArgGlyArgGlnLys 620
QY 2062 ATCGTGAGCTGACCGAGACCACCAAGACAGACCGAGGTGCGAGGCCATCCAGCTGGCC 2121
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCAGACAGCGCGCAGCGAGGTGAACATCGTCACCGACACAGCGGTGCGGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIle 660
QY 2182 ATCCAGGCCAGCCCGACAGAGCGAGAGCGAGCTGGTGAACCATCATCATCGAGCGCTG 2241
Db 661 IleGlnAlaGlnProaspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAGAGAGGTGTACCTGAGCTGGTGGTGGCCGCCACAGGGGCATCGGGGGCAAC 2301
Db 681 IleLysLysGlnLysValTyLeuThrTrpIleProAlaHisLysGlyIleGlyGlyAsn 700
QY 2302 GAGCAGATCAGACAGCTGGTGAGCAAGGGGCATCCGCAAGGTGCTGTTCTTGGACGGCATC 2361
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db 721 Glu 721
```

## RESULT 9

```
US-10-325-468-9
; Sequence 9, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;
; OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;
; OTHER INFORMATION: clone P10.26 protein Pol
US-10-325-468-9
```

## Alignment Scores:

Pred. No.:	1.38e-158	Length:	1003
Score:	3427.00	Matches:	638
Percent Similarity:	94.59%	Conservative:	44
Best Local Similarity:	88.49%	Mismatches:	33
Query Match:	74.94%	Indels:	6
DB:	16	Gaps:	3

US-09-610-313B-31 (1-2463) x US-10-325-468-9 (1-1003)

```
QY 220 TTCTTCGGCAGGACCTGGCTTCCCGCAGGCGAGCGCGGAGTTCCTCCCGAGGAGCC 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
```

```
QY 280 AACCGCGCAACAGCCCGCCACCGCCGAGCTGCAGTG-----CGCGCGCAACAACCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCCAGGCGCACCCCTG-----AACTTCCCGCAGATCAC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCGGCAGATCAAGGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCAGCACACCGCTGTGTGAGAGATGAGCTGCCCGCAAGTGAAGACCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
QY 508 AGATGATCGGCGGATCGCGGCTTCATCAAGGTGGCGCAGTACGACGATCCTGTATC 567
Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTyArgGlnIleProIle 120
QY 568 GAGATCTCGGGCAAGAGCCATCGGCACCGCTGTGTGATCGGCCCGCCACCGCGTGAACATC 627
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGCTGCACCTGAACCTTCCCATCAGCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGAACTGAAGCCCGGCATCGAGCGGCCCGCCCAAGGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAAAGAGGGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGGCCCGGAGAACCCCTACAACACCCCGCTGTTCGCATCAAGAAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAGCTGTGTGACTTCGCGAGCTGAACAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240
QY 928 GACTTCTGGAGGTGCAGCTGGGCATCCCGCCACCGCGCGCTGAAGAGAAAGAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGACGTGGGCGACCGCTACTTACGGTGGCCCTTGGACGAGGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTACCATCCCGCAGCATCAACAGGAGACCCCGCGCATCCGCTACCG 1107
Db 281 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 300
QY 1108 TACAACGTGTGCGCCCGCAGGCTCGAAGGGCAGCCCGCAGCATCTTCAGAGCAGCATGACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTGTGAGCCCTTCCCGCGCGCAACCCCGAGATCGTGATCTACCGAG-----GCC 1221
Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyGlnTyMetAsp 340
QY 1222 CCCCTGTAGTGGGCGCAGCAGCTGGAGATCGGCAGCACCGCCCGCAAGATCGAGGAGCTG 1281
Db 341 AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGTGTGGCTGGGCTTCCACCCACCGCGCACAAGAGCAACCAAGAGGAGCC 1341
Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTTCTGTGGATGGGCTACGAGCTGACCGCCCGCACAAGTGGACCGGTGACCCATCGAG 1401
```

```

Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1402 CTGCCGAGAGGAGAGCTGACCGCTGAACACATCCAGAACCTGGTGGGCAAGCTGAAC 1461
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1462 TGGCCAGCAGATCTACCCCGGCGATCAAGTGGCGCAGCTGTCAAGCTGTCTGCGCGGC 1521
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
QY 1522 GCCAAGSCCTGACCGACATGTCGCCCTGACCGAGGAGCCGAGCTGGAGCTGGCCGAG 1581
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluAlaGlu 460
QY 1582 AACCGCAGATCTCTGCGAGCGCGCTGACGGCTGTACTACGACCCCGACGAGCCTG 1641
Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
QY 1642 GTGCCGAGATCCAGAGAGGCGCCACGACGATGGACCTACCAAGTCTACCAAGAGCC 1701
Db 481 IleValGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIlePheGlnGluPro 500
QY 1702 TTCAGAACCTGAGACCGGACGACGACGACGACGACGACGACGACGACGACGACG 1761
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
QY 1762 AAGCAGCTGACCGAGGCGCTGCAGAGATCCGATGAGAGCATCGATGCTGGGCAAG 1821
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
QY 1822 ACCCCAAAGTTCGCCCTGCCCTCCAGAGGAGACCTGGGAGACCTGGTGGACCGATC 1881
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr 560
QY 1882 TGGCAGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG 1941
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1942 TCGTACAGCTGAGAGAGGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAGCGGCC 2001
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 2002 GCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGAGCGCGCGCAGAAG 2061
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
QY 2062 ATCGTAGCCTGACCGAGACCAACCAAGACCGAGCTGCAGGCCATCCAGCTGGCC 2121
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2122 CTGCAGGACGCGCAGCAGGAGTGAACATCGTACCGACGACGAGCAGTACGCCCTGGGCATC 2181
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2182 ATCCAGCCCGCAGCCGCAAGAGCAGGAGCTGGTGAACAGATCATCGACGAGCTG 2241
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2242 ATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCCCGCCACAGGGCATCGCGCGCAAC 2301
Db 681 IleLysLysGlnLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
QY 2302 GAGCAGATCCACAGCTGGTGGAGCAGGCGCATCCGCAAGGTGCTGTCCTGGAGCGGCATC 2361
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720
QY 2362 GAT 2364
Db 721 Glu 721

```

RESULT 10  
US-10-325-468-20

```

; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20

```

```

Alignment Scores:
Pred. No.: 1,38e-158 Length: 1003
Score: 3427.00 Matches: 638
Percent Similarity: 94.59% Conservative: 44
Best Local Similarity: 88.49% Mismatches: 33
Query Match: 74.94% Indels: 6
DB: 16 Gaps: 3

```

US-09-610-313B-31 (1-2463) x US-10-325-468-20 (1-1003)

```

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
QY 280 AACCGCGCCAAACGCCCGCCAGCCAGCGAGCTGCAGGTG-----CCGCGCGCAACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
QY 508 AACATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
Db 101 LysMetIleGlyIleGlyIleGlyIleLysValArgGlnTyrAspGlnIleProIle 120
QY 568 GAGATCTCGCGCAAGAGCGCATCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGGTGCGACCTGCACTTCCCATCAGCCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGAGCTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200

```





Db 21 ThrGlyAlaAsnSerSerAlaSerArgLysLeu-----GlyAspGly----- 34  
QY 340 GAGCGCGCGCGAGCCAGGCGCACC-----CTGAATCTCCCCAGATCACCCCTG 390  
Db 35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53  
QY 391 TGGCAGCGCCCTGTGTGAGCATCAAGGTGGGGCCAGATCAAGAGAGCCCTGTCTGAC 450  
Db 54 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 73  
QY 451 ACCGGCGCCACGACACCGGTCTGGAGGAGATGAGCTGCGCCCGCAAGTGGAGCCCAAG 510  
Db 74 ThrGlyAlaAspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLys 93  
QY 511 ATGATCGCGCGCATCGCGCGCTTCAATCAAGGTGGCCAGTACGACACCATGATCGAG 570  
Db 94 MetIleGlyGlyIleGlyGlyPheIleLysValArgGlnThrAspGlnIleLeuIleGlu 113  
QY 571 ATCTGCGGCAAGAGCGCATCGGCACCGTGTGATCGCGCCCGCCCGTGAACATCATC 630  
Db 114 IleCysGlyLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133  
QY 631 GCGCGCAACATGTCACCGCAGCTGGCTGGCTGCACCTGAATCTCCCATCAGCCCATCGAG 690  
Db 134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153  
QY 691 ACCGTGCGCTGAAGCTGAAGCCCGCATGACCGCGCCCAAGGTGAAGCAGTGGCCCTG 750  
Db 154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 173  
QY 751 ACCGAGGAGAGATCAAGCGCCCTGACCGCCATCTGCGAGGAGATGGAAGAGGGCAAG 810  
Db 174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGlyLys 193  
QY 811 ATACCAAGATCGCCCGGAGAACCTCTACACACCGCCGTGTGCGCATCAAGAAGAG 870  
Db 194 IleSerLysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLys 213  
QY 871 GACAGCACCAAGTGGCGCAAGCTGTGACTTCCCGCAGCTGAACAAGCGCACCCAGGAC 930  
Db 214 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233  
QY 931 TTCTGGAGGTGACAGCTGGGCGATCCCCACCGCCCGCCCTGGAAGAAAGAGAGCGTG 990  
Db 234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 253  
QY 991 ACCGTGTGAGCTGGGCGACGCTACTTCAAGGTGCGCTGGAGGAGACTTCCGCGAAG 1050  
Db 254 ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluSerPheArgLys 273  
QY 1051 TACACCGCTTCAACATCCCGCAGCATCAACAAGAGACCCCGCATCCGCTACAGTAC 1110  
Db 274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTrp 293  
QY 1111 AACGTGTGCGCCAGGCTGGAAGGGCAGCGCCAGCATCTTCCAGAGCAGCATGACCAAG 1170  
Db 294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313  
QY 1171 ATCTGAGGCTTCCCGCGCCCGCAACCGCGAGATCGTATCTACCAAG-----GCCGCC 1224  
Db 314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyGlnTrpMetAspAsp 333  
QY 1225 CTGTACGTGGCGACGACCTGGAGATCGCGCAGCAGCGCCAGATCGAGGAGCTCGCG 1284  
Db 334 LeuTyThrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353  
QY 1285 AAGCACCTGTCTGGCGCTGGGCTTCAACACCGCCCGCAGAGACACCAAGAGAGGAGCCGCC 1344  
Db 354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373  
QY 1345 TTCTGTGGATGGCTACGAGCTGCACCGCGACAAAGTGGACCGGTGACCGCATCGAGCTG 1404

Db 374 PheLeuTrpMetGlyTyThrGluLeuHisProAspArgTrpThrValGlnProIleGluLeu 393  
QY 1405 CCCGAGAAGGAGAGCTGGACCGCTGAACGACATCCAGAAAGCTGTGGGCAAGCTGAACCTGG 1464  
Db 394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413  
QY 1465 GCAGAGCAGATCTACCCCGGCATCAAGGTGGCGCAGCTGTGCAAGCTGTGCGCGGCGCC 1524  
Db 414 AlaSerGlnIleTyAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyThr 433  
QY 1525 AAGSCCTCTGACCGACATCGTGGCCCTGACGAGGAGCGCGAGCTGGAGCTGGCGAGAAC 1584  
Db 434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluGluAsn 453  
QY 1585 CGGAGATCTCTGGCGAGCGCGCTGACGCGGTGTACTACGACCCCGACAGGAGCTGGTG 1644  
Db 454 ArgGluIleLeuArgGluProValHisGlyValTyThrAspProSerLysAspLeuVal 473  
QY 1645 GCGAGATCCAGAGAGGCGCCAGCAGCTGGACCTACAGATCTACAGGAGCCCTTC 1704  
Db 474 AlaGluValGlnLysGlnGlyAspGlnTrpThrTyThrGlnIleTyThrGlnGluProPhe 493  
QY 1705 AAGAACCTGAAGACCGCAAGTACGCAAGATGCGCAAGCGCCGACCAACAGCTGAAG 1764  
Db 494 LysAsnLeuLysThrGlyLysTyThrSerArgLysArgSerAlaHisThrAsnAspValArg 513  
QY 1765 CAGCTGACCGAGCGCTGCAAGAAGATCGCCATGCGAGAGCATCTGTGATCTGGGCAAGACC 1824  
Db 514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533  
QY 1825 CCAGAGTTCGCTGCGCATCCAGAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1884  
Db 534 ProlLysPheArgLeuProIleGlnArgGluThrTrpGluThrTrpTrpMetGluTyTrp 553  
QY 1885 CAGGCACCTGGATCCCGAGTGGGAGTCTGTGAACACCCCGCCCTGGTGAAGCTGG 1944  
Db 554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeuTrp 573  
QY 1945 TACAGCTGAGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCCGCC 2004  
Db 574 TyrGlnLeuGluLysAspProIleValGlyAlaGluThrPheTyThrValAspGlyAlaAla 593  
QY 2005 AACCGGAGACCAAGATCGGAGCGCGCTAGTGAACCGCGCGCGCGCGCGAGAAATC 2064  
Db 594 SerArgGluThrLysLeuGlyLysAlaGlyTyThrValThrAspArgGlyArgGlnLysVal 613  
QY 2065 GTGAGCTGACCGGAGCCACCAAGACCGAGCTGCAGGCGCATCCAGCTGGCGCTG 2124  
Db 614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
QY 2125 CAGGACAGCGCGCAGCGAGGTGAACATCGTGAACCGACAGCAGCTACGCGCTGGGCATCATC 2184  
Db 634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyThrAlaLeuGlyIleIle 653  
QY 2185 CAGCGCCAGCCGCAAGAGCAGGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATC 2244  
Db 654 GlnAlaGlnProAspArgSerGluValValSerGlnIleIleGluGluLeuIle 673  
QY 2245 AAGAGGAGAGAGCTGTACCTGAGCTGGTGGTCCCGCCACCAAGGCGCATCGCGCGCAAGCAG 2304  
Db 674 LysLysGlnLysValTyThrLeuSerTrpValProAlaHisLysGlyIleGlyLysGlu 693  
QY 2305 CAGATCGACAGCTGGTGTGACCAAGGCGCATCGCAAGGTGTCTTCTGGAGCGCATCGAT 2364  
Db 694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713

## RESULT 12

US-10-296-734-1470  
; Sequence 1470, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A



```

Db      594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613
Qy      2065 GTGAGCCTGCAGCAGACCAACCAAGAGCCAGAGTGCAGGCGCATCCAGCTGGCCCTG 2124
Db      614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633
Qy      2125 CAGACACGCGGCAGCAGAGGTGAACATCGTGAACCGACGACCCAGTACGGCCCTGGGGCATCATC 2184
Db      634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653
Qy      2185 CAGGCCCAGCCCGCAAGAGCGAGCGAGCGAGCTGGTGAACACGATCATCGAGCAGCTCATC 2244
Db      654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleIleGluIleIle 673
Qy      2245 AAGAAGGAGAGAGGTGTACTCTGAGCTGGGTGCCCGCCCAAGAGGCATCGGCGGCAACGAG 2304
Db      674 LysLysGluLysValTyrLeuSerTyrProAlaHisLysGlyIleGlyGlyAsnGlu 693
Qy      2305 CAGATCAACAAGCTGGTGAGCAAGGGGATCGCGAAGGTGCTGTTCTTCTGGACGGCATCGAT 2364
Db      694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713

RESULT 13
US-10-296-734-1471
; Sequence 1471, Application US/10296734
; Publication No. US20040054137A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Scott A
; APPLICANT: Ramshaw, Ian A
; TITLE OF INVENTION: Synthetic molecules and uses therefor
; FILE REFERENCE: Savine
; CURRENT APPLICATION NUMBER: US/10/296,734
; PRIOR FILING DATE: 2003-08-04
; PRIOR APPLICATION NUMBER: AU PQ7761/00
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 1507
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1471
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: POL Consensus A
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (15)..(15)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (22)..(23)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (26)..(26)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (30)..(30)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (36)..(36)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (38)..(39)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (41)..(44)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE

```

```

; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (335)..(335)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (365)..(365)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (369)..(369)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (373)..(373)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(403)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (434)..(434)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (439)..(439)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (452)..(452)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (470)..(470)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (480)..(480)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (487)..(487)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (514)..(514)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (534)..(535)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (544)..(544)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (553)..(553)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (593)..(593)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (610)..(610)
; OTHER INFORMATION: X = unknown amino acid

```

```

; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (675)..(675)
; OTHER INFORMATION: X = unknown amino acid
; FEATURE:

Alignment Scores:
Pred. No.:      4.62e-150      Length:      1006
Score:          3251.50        Matches:      618
Percent Similarity: 88.81%      Conservative: 25
Best Local Similarity: 85.36%    Mismatches:   72
Query Match:     71.10%        Indels:       9
DB:              15           Gaps:         3

US-09-610-313B-31 (1-2463) x US-10-296-734-1471 (1-1006)
QY 220 TTCTTCGCGAGGACCTGGCTTCCCTCCAGGGAAGCCGCGAGTTCCTCCAGCGAGCAG 279
    |||||:::|||||
Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArg***PheSerSerGluGln 20
QY 280 AACCGCGCCAAACAGCCCAACAGCCGCGAGCTGCGGCGGCGACAAACCC----- 333
    |||||
Db 21 Thr*****AsnSer***ThrSerArg***LeuTrpAspGlyGly***Asp*****Leu 40
QY 334 CGCAGCGAGGCGCGCGCGCGAGCGCCAGGCGC-----ACCTGTGAATTCCTCCC 378
    |||||
Db 41 *****Gly***Glu***GlnGly***Gly*****SerPhe***PhePro 60
QY 379 CAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGGTGGCGGCGCAGATCAAGGAG 438
    |||||
Db 61 GlnIleThrLeuTrpGlnArgProLeuValThrVal***Ile***GlyGlnLeuIleGlu 80
QY 439 GCGCTGTGGACACCGCGCGCGACGACCGCTGCTGGAGGAGATGAGCTGCCCGGCAAG 498
    |||||
Db 81 AlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAsnProGlyLys 100
QY 499 TGAAGCCCAAGATGATCGCGGCATCGCGCGCTTCATCAAGGTGGCGGCGCAGTACGACCAG 558
    |||||
Db 101 TrpLysProLys***IleGlyGlyIleGlyGlyPheIleLysValArgGlnTrpAspGln 120
QY 559 ATCTGTATCGAGATTCGCGCAAGAGCCATCGGCACCGTGTGTGATCGGCCCGCCACCC 618
    |||||
Db 121 IleLeuIleGluIleCysGlyLysLys***IleGlyThrValLeuValGlyProThrPro 140
QY 619 GTGAACATCATCGCGCGCAACATGCTGACCCCGTGGCTGCGACCTGCACTTCCCATC 678
    |||||
Db 141 ValAsnIleIleGlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160
QY 679 AGCCCATCGAGACCCGTCGCGTGAAGCTGAAGCCCGCATGAGCGCGCGCGCGCGCGCG 738
    |||||
Db 161 SerProIleGluThrValProValLysLeuLysPro***MetAspGlyProLysValLys 180
QY 739 CAGTGGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGCCCATCTGCGGAGGAGATGGAG 798
    |||||
Db 181 GlnTrpProLeuThrGluGluLysIleLysAlaLeuThr***IleCys***GluMetGlu 200
QY 799 AGGAGGCGCAAGATCACCAAGATCGCGCGCGAGAACCCCTACACACCCCGCTTCCGCC 858
    |||||
Db 201 LysGluGlyLysIleSerLysIleGlyProGluAsnProTrpAsnThrProValPheAla 220
QY 859 ATCAAGAAGAAGGACAGACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAG 918
    |||||
Db 221 IleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLys 240
QY 919 CGCACCCAGGACTTCTGGGAGGTGCGAGCTGGGGCATCCCCACCCCGCGCGCTGAAGAAG 978
    |||||
Db 241 ArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis***AlaGlyLeuLysLys 260
QY 979 AAGAAGCGGTGACCGCTGCTGGGCGCGCGCTTACTTTCAGCTGCGCTTGGAGCAG 1038
    |||||
Db 261 LysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAsp*** 280
QY 1039 GACTTCCGCAAGTACACCGCCTTCCACCATCCCGCAGCATCAACACGAGACCCCGCGCATC 1098

```

Db 281 \*\*\*PheArgLysTyrThrAlaPheThrIleProSer\*\*\*AsnAsnGluThrProGly\*\*\* 300  
QY 1099 CGCTACCAAGTACAAAGCTGCTCCCGAGGGCTGGAGGCGACCCCGACGATCTTCCAGAGC 1158  
Db 301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro\*\*\*IlePheGln\*\*\* 320  
QY 1159 AGCATGACCAAGATCTCTGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCTACAG 1218  
Db 321 SerMetThrLysIleLeuGluProPheArg\*\*\*\*\*Pro\*\*\*IleValIleTyrGln 340  
QY 1219 -----GCCCCCTGTACGTGGGCGAGCGACTCGAGATCGCCAGCAGCCGCGCAAGATC 1272  
Db 341 TyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360  
QY 1273 GAGGAGCTGGCGCAAGCACTGCTCGCTGGGGCTTCCACACCCCGACCAAGAACACAGC 1332  
Db 361 GluGluLeuArg\*\*\*HisLeuLeu\*\*\*TrpGlyPhe\*\*\*ThrProAspLysLysHisGln 380  
QY 1333 AAGGAGCCCTTCTGTGTGATGGGTACAGCTGCACCCCGACAAAGTGACCGTGCAG 1392  
Db 381 LysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThrValGln 400  
QY 1393 CCCATCGAGTGGCGGAGAGGAGCTGGACCTGACCGATCGACACATCCAGAGCTGGTGGC 1452  
Db 401 Pro\*\*\*\*\*LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420  
QY 1453 AAGCTGAACCTGGGCGAGCATCTACCCCGCATCAAGGTGGCCAGCTGTGCAAGCTG 1512  
Db 421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys\*\*\*LysGlnLeuCys\*\*\*Leu 440  
QY 1513 CTGCGCGCGCCCAAGGCCCTGACCGACATCGTGCCTCCCTGACCGAGGCGCGAGCTGAG 1572  
Db 441 LeuArgGlyAlaLysAlaLeuThrAspIleVal\*\*\*LeuThrGluGluAlaGluLeuGlu 460  
QY 1573 CTGCGCGAGAACCGGAGATCTCGCGGAGCCCTGCAGCGCTGTACTAGACCCCGCAGC 1632  
Db 461 LeuAlaGluAsnArgGluIleLeuLys\*\*\*ProValHisGlyValTyrTyrAspPro\*\*\* 480  
QY 1633 AAGGACCTGTGGCGGAGATCCAGAAGCAGGCGCCACACAGTGGACCTACCATCTAC 1692  
Db 481 LysAspLeuValAlaGlu\*\*\*GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500  
QY 1693 CAGGAGCCCTTCAAGAACCTGAAGACCGGAGTACGCCAAGATGCGCACCGCCACAC 1752  
Db 501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla\*\*\*LysArgSerAlaHisThr 520  
QY 1753 ACGACGTGAGCAGCTGACGAGCGCGTGCAGAAATCGCCATGCGGAGAGCATCGTGATC 1812  
Db 521 AsnAspValLysGlnLeuThrGluValValGlnLysVal\*\*\*\*\*GluSerIleValIle 540  
QY 1813 TGGGCGAAGACCCCAAGTTCGCGCTGCCATCCAGAAGGAGACCTGGGAGACCTGGTGG 1872  
Db 541 TrpGlyLys\*\*\*ProLysPheArgLeuProIleGln\*\*\*GluThrTrpGlu\*\*\*TrpTrp 560  
QY 1873 ACCGACTACTGCGAGGCGACCTGTGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTG 1932  
Db 561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580  
QY 1933 GTGAAGCTGTGTACCAAGCTGGAGAGAGGCCCATCATCTCGCGCGCGAGACCTTCTAGTG 1992  
Db 581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle\*\*\*GlyAlaGluThrPheTyrVal 600  
QY 1993 GACGCGCGCCCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTGACCGCGCGGCG 2052  
Db 601 AspGlyAlaAlaAsnArgGluThrLys\*\*\*GlyLysAlaGlyTyrValThrAspArgGly 620  
QY 2053 CGGCAAGATCTGTGAGCTGACCGACACCAACCAAGAGACCGAGCTGCAGCGCCATC 2112  
Db 621 ArgGlnLysValValSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIle 640  
QY 2113 CAGCTGCGCTGCAAGGACAGCGGAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCC 2172

Db 641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660  
QY 2173 CTGGGCAATCATCCAGGCGCCAGCCGACAGAGCGAGAGCGAGCTGGTGAACACGATCATC 2232  
Db 661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu\*\*\*ValAsnGlnIleIle 680  
QY 2233 GAGCAGCTGATCAAGAAGGAGAGGTGTACTCTGAGCTGGGTGGCCGCCCAAGAGGCGATC 2292  
Db 681 GluLysLeuIle\*\*\*Lys\*\*\*LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700  
QY 2293 GGCGGCAACGAGCAGATCGACAGCTGTGAGCAAGGCGCATCCGCAAGGTCTGTCTCTG 2352  
Db 701 GlyGlyAsnGluGlnValAspLysLeuValSer\*\*\*GlyIleArgLysValLeuPheLeu 720  
QY 2353 GACGCGCATCGAT 2364  
Db 721 AspGlyIleAsp 724  
RESULT 14  
US-10-301-661A-6  
; Sequence 6, Application US/10301661A  
; Publication No. US20030157660A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-  
; APPLICANT: INSERM  
; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: MAUCLERE, Philippe  
; APPLICANT: LOUSSERT-AJAKA, Ibtissem  
; APPLICANT: SIMON, Francois  
; APPLICANT: SARAGOSTI, Sentob  
; APPLICANT: BARRE-SINOSSI, Francoise  
; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.  
; FILE REFERENCE: 598US12  
; CURRENT APPLICATION NUMBER: US/10/301,661A  
; CURRENT FILING DATE: 2002-11-22  
; PRIOR APPLICATION NUMBER: US/09/319,588C  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: FR96/15087  
; PRIOR FILING DATE: 1996-12-09  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 1014  
; TYPE: PRT  
; ORGANISM: Human immunodeficiency virus type 1  
US-10-301-661A-6  
Alignment Scores:  
Pred. No.: 9.3e-147 Length: 1014  
Score: 3183.50 Matches: 583  
Percent Similarity: 90.25% Conservative: 74  
Best Local Similarity: 80.08% Mismatches: 58  
Query Match: 69.62% Indels: 13  
Gaps: 4  
US-09-610-313B-31 (1-2463) x US-10-301-661A-6 (1-1014)  
QY 220 TTCTTCGCGAGGACCTGGCTTCCCGCCAGGCGAGCCCGCGAGTTCCTCCAGCGAGCAG 279  
Db 1 PhePheArgGluGluLeuValSerLeuGlnArgGluThrArgLysLeuProAspAsn 20  
QY 280 AAC-----CGCGCAACAGCCCGCCAGCGCGCGAGCTGCAGGTG----- 318  
Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGluGlu 40  
QY 319 ---CGCGGCGCAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366  
Db 41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60  
QY 367 CTGAACCTTCCCGCAGATCACCCTGTGCGAGCGCCCTGTGAGCATCAAGGTGGCGGC 426  
Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80

QY	427	CAGATCAAGGAGGCCCTGTCTGGACACCGGCGCCGACACACCGTGTCTGGAGGAGATGACG	486
DB	81	GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluLeuLeuGln	100
QY	487	CTGCGCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGCGCTTCATCAAGGTGCCG	546
DB	101	LeuGluGlyLysTrpLysProLysMetIleGlyIleGlyIleGlyPheIleLysValArg	120
QY	547	CAGTACGACCAGATCCTGATCGAGATCTGGCGCAAGAAGCCATCGCAGCCGTGCTGATC	606
DB	121	GlnTrpAspAsnIleThrValAspIleGlnGlyArgAsnLeuLeuValGlyThrValLeuVal	140
QY	607	GGCGCCACCCCGTGAACATCATCTGGCGCGCAACATGCTGACCCAGCTGGGCTGCACCCCTG	666
DB	141	GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu	160
QY	667	AATTTCCCATACGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGGC	726
DB	161	AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly	180
QY	727	CCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAGATCAAGGCCCTGACCGCCACTCTGC	786
DB	181	ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys	200
QY	787	GAGGAGATGAGAAAGGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTTACAACACC	846
DB	201	ThrGluMetGluLysGluGlyLysIleSerArgIleGlyProGluAsnProTyrAsnThr	220
QY	847	CCCGTGTTCCTCCATCAAGAAAGAGCAGACCAAGTGGCGCAAGCTGTGTGACTTCCCG	906
DB	221	ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg	240
QY	907	GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCCGCACCCCGCC	966
DB	241	GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla	260
QY	967	GGCTGAAGAAAGAGAGACGCTGACCGTCTGACGTGGCGCAGCGCTACTTTCAGCGTG	1026
DB	261	GlyLeuLysGlnLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys	280
QY	1027	CCCCTGACGAGGACTTCCGCAAGTACACCGCCTTCACCATCCCGACATCAACACGAG	1086
DB	281	ProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGlu	300
QY	1087	ACCCCGGCACTCGCTACAGTACACGCTGCTGCCCGCCAGGCTGGAAGGCGAGCCCGCAGC	1146
DB	301	ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla	320
QY	1147	ATCTTCCAGACGACATGACCAAGATCCTCGAGCCCTTCGCGCGCCCAACCCCGAGATC	1206
DB	321	IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysHisProGluIle	340
QY	1207	GTGATCTAACAG-----GCCCGCCTGTAGCTGGGACGACCTGGAGATCGCGCAGCAC	1260
DB	341	IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis	360
QY	1261	CGCGCCAAAGATCGAGGAGCTCGCAAGCACTGCTGCGCTGGGCTTTCACACCCCGCAGC	1320
DB	361	ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp	380
QY	1321	AAGNAGCACCAAGAGGACCCCTTCTGTGGATGGCTACGAGTGCACCCCGACAAG	1380
DB	381	LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys	400
QY	1381	TGGACCTGACGCCCATCGAGTCCCGAGAAGAGAGCTGGACCGGTGAACGACATCCAG	1440
DB	401	TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln	420
QY	1441	AAGCTGTGGGCAAGCTGAAGTGGCGCAGCCAGATCTACCCCGGCATCAAGGTGCGCCAG	1500
DB	421	LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln	440

Qy	1501	CTGTGCAAGCTGCTGGCGCGCGCAAGGCCCTGACCGCATCGTGTGCCCTCGACCGGAGG	1560
Db	441	LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu	460
Qy	1561	GCCGAGCTGGAGCTGGCCGAGAACCCGAGATCCTGCGCGAGGCCGTGACGGCGGTGAC	1620
Db	461	AlaGluLeuGluLeuAlaGluAsnArgGluLeuLeuLysGluProLeuHisGlyValTyr	480
Qy	1621	TACGACCCCAAGCAAGCACTGTGTGGCGAGATCCAGAAGCAGGGCCACGACGAGTGACC	1680
Db	481	TyrAspProGlyLysGluLeuValAlaGluLeuGlnLysGlnGlyGlnGlyGlnTyrThr	500
Qy	1681	TACGAGATTCACGAGAGCCCTTCAAGAAGCTGAAGACCGGCAAGTACGCCAAGATGCCG	1740
Db	501	TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg	520
Qy	1741	ACGCGCCCAACCAACGAGCTGAAGCAGCTACCGAGCGCGCTGCAGAAAGATCGCATGGAG	1800
Db	521	SerAlaHisThrAsnAspIleLysGlnLeuValGluValArgLysValAlaThrGlu	540
Qy	1801	AGCATCGTCACTGGGGCAAGACCCCAAGTTCCGCTGCCCCTCAGAGAGGAGACCTGG	1860
Db	541	SerIleValIleTyrGlyLysThrProLysPheArgLeuProValGlnLysGluValTyr	560
Qy	1861	GAGACCTGTGTGACCACTACCTGGCAGGCCACCTGATCCCGAGTGGGAGTTCGTGAAC	1920
Db	561	GluAlaTyrTyrThrAspHisTyrGlnAlaThrTyrIleProGluTyrGluPheValAsn	580
Qy	1921	ACCCCGCCCTGTGTGAGCTGTGTACCAAGCTGGAGAAGGCCCATCATCGCGCGCAG	1980
Db	581	ThrProLeuValLysLeuTyrTyrGlnLeuGluThrGluProIleSerGlyAlaGlu	600
Qy	1981	ACCTTCTACTGTGACCGCGCGCAACCCGCGAGACCAAGATCGGCAAGCCGCTACGTG	2040
Db	601	ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal	620
Qy	2041	ACCGACCGGCGCGCAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGACCGAG	2100
Db	621	ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu	640
Qy	2101	CTGCAGGCCATCAGCTGGCCCTGCGAGGACGCGGCGAGGAGTGAACATCGTGACCGAC	2160
Db	641	LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp	660
Qy	2161	AGCCAGTACGCCCTGGCGCATCATCGAGGCCCGCCGCAAGAGCAGAGCGAGCTGGTG	2220
Db	661	SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal	680
Qy	2221	AACGAGATCATCAGCAGCTGTATCAAGAAGGAGAGGTGTACTGAGCTGGGTGCCGCC	2280
Db	681	SerGlnIleIleGluGluLeuIleLysLysGluArgValTyrLeuSerTyrValProAla	700
Qy	2281	CACAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGACGAGGGCATCCGCAG	2340
Db	701	HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleThrGly	720
Qy	2341	GTGCTGTTCCTGACCGCATCGAT	2364
Db	721	IleLeuPheLeuAspGlyIleGlu	728
RESULT 15			
US-09-952-060-35			
; Sequence 35, Application US/09952060			
; Publication No. US2003004421A1			
; GENERAL INFORMATION:			
; APPLICANT: Emini, Emilio A.			
; APPLICANT: Youil, Rima			
; APPLICANT: Bett, Andrew J.			
; APPLICANT: Chen, Ling			
; APPLICANT: Kaslow, David C.			
; APPLICANT: Shiver, John W.			
; APPLICANT: Toner, Timothy J.			
; APPLICANT: Casimiro, Danilo R.			



; TITLE OF INVENTION: ENHANCED FIRST GENERATION ADENOVIRUS  
; TITLE OF INVENTION: VACCINES EXPRESSING CODON OPTIMIZED HIV1-GAG, POL, NEF AND  
; FILE REFERENCE: 20747Y  
; CURRENT APPLICATION NUMBER: US/09/952,060  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: PCT/US01/28861  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/317,814  
; PRIOR FILING DATE: 2001-09-07  
; PRIOR APPLICATION NUMBER: 60/279,056  
; PRIOR FILING DATE: 2001-03-27  
; PRIOR APPLICATION NUMBER: 60/233,180  
; PRIOR FILING DATE: 2000-09-15  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 1350  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Codon optimized gag-1A pol fusion  
US-09-952-060-35

Alignment Scores:  
Pred. No.: 1.22e-143 Length: 1350  
Score: 3119.50 Matches: 599  
Percent Similarity: 81.16% Conservative: 43  
Best Local Similarity: 75.73% Mismatches: 59  
Query Match: 68.22% Indels: 90  
DB: 10 Gaps: 6

US-09-610-313B-31 (1-2463) x US-09-952-060-35 (1-1350)

QY	13	ATGCGCGAGCCATGACCGAGCCACC---AGCGCAACATCTGTATGACGCGCAGCAAC	69
DB	363	LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn	382
QY	70	TTCAAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGCGCAAGGAGGCGCCACATCGCC	129
DB	383	PheArgAsnGlnArgLysThrValLysCysPheAsnCysGlyLysValGlyHisIleAla	402
QY	130	CGCAACTGCGCGCCCCCGCAGAGAGGGTGTCTGGAAGTGGCGCAAGAGGGCCACCAG	189
DB	403	LysAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGln	422
QY	190	ATCAAGGACTGCACCGAGCGCCAGGCAACTTCTCCGCGAGGACCTGGCTTCCCCCAG	249
DB	423	MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyLysIleTrpProSerHisLys	442
QY	250	GGCAAGGCGCGAGTTCCCGAGCGAGAGCAACCGCGCCACAGCCCCACAGCCGCGAG	309
DB	442	SGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaProProGluGluSe	462
QY	310	CTGAGGTGGCGG-----CGGCAACCCCCCGAGCGAGCGCGCGCGCGCGCGAGGC	363
DB	462	rPheArgPheGlyGluGluLysThrProSerGlnLysGlnGluProIleAspLys--	481
QY	364	ACCTGTGAAGTCTCCCGCAGATCACCTCTGGCAGCGCCCTGTGTGAGCATCAAGTGGGC	423
DB	482	GluLeu-----TyroProLeuAlaSerLeuArg-----	490
QY	424	GGCCAGATCAAGGAGGCGCTGCTGGACACCGCGCGCGAGCACACCGTGTGGAGGAGATG	483
DB	490	-----	490
QY	484	AGCCTGCCCGCAAGTGGAGAGCCCAAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTG	543
DB	490	-----	490
QY	544	CGCCAGTACGACGATCTCTGATCGAGATCTCGCGGAGAGGCCATCGGCACCGTGTCTG	603
DB	490	-----	490

QY	604	ATCGGCCCCACCCCGGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGGTGCACC	663
DB	491	-----SerLeuPheGlyAsnAspProSerSerGlnMetAla-----	502
QY	664	CTGAACCTTCCCATCGAGCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCGGAC	723
DB	503	-----ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAsp	519
QY	724	GGCCCCAAGGTGAAGCAGTGGCGCCCTGACCGAGAGAGATCAAGGCGCTGACCGCCATC	783
DB	520	GlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIle	539
QY	784	TGCAGGAGATGAGAGAGAGGCAAGATCACCAAGATCGGCCCCCGAGAACCCCTCAAC	843
DB	540	CysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyAsn	559
QY	844	ACCCCGCTGTTCGCCATCAAGAGAGAGGACAGCAGCAGGAGTGGCGCAGCTGGTGGACTC	903
DB	560	ThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPhe	579
QY	904	CGCGAGCTGAACCAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGCGCATCCCCACCCC	963
DB	580	ArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPro	599
QY	964	GCGGCTGAAGAGAGAGCGTGACCGTGTGACGTGGCGAGCGCTACTTCAGC	1023
DB	600	AlaGlyLeuLysLysLysSerValThrValLeuAlaValGlyAspAlaTyPheSer	619
QY	1024	GTGCCCCGTGACGAGGACTTCCCGAAGTACACCGCTTCCACCATCCCGAGCATCAAC	1083
DB	620	ValProLeuAspGluAspPheArgLysTyThrAlaPheThrIleProSerIleAsnAsn	639
QY	1084	GAGACCCCCGGCATCCGCTACCAAGTCAACGTGTGCCCGAGGCTGGAGGGGAGCGCC	1143
DB	640	GluThrProGlyIleArgTyThrGlnTyAsnValLeuProGlnGlyTrpLysGlySerPro	659
QY	1144	AGCATCTTCAGAGCAGCATGACCAAGATCCTCGAGCCCTTCCGCGCCCGCAACCCGAG	1203
DB	660	AlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAsp	679
QY	1204	ATCGTGTATACAG-----GCCCGCTGTAGTGGCGAGCGACCTGGAGATCGGCAG	1257
DB	680	IleValIleTyThrGlnTyMetAlaAlaLeuTyValGlySerAspLeuGluIleGlyGln	699
QY	1258	CACCGCGCAAGATCCAGGAGCTGCGCAAGCACCTGTGCGCTGGGCTTCCACCCACCCC	1317
DB	700	HisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrPro	719
QY	1318	GACAAGACACAGAGGAGCGCCCTTCTGTGATGGCTACGAGCTGCAACCCGAC	1377
DB	720	AspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyThrGluLeuHisProAsp	739
QY	1378	AAGTGCAGCTGAGCCCATCGAGCTGCGCGAGAGAGAGAGCTGGACCGTGAACGATC	1437
DB	740	LysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIle	759
QY	1438	CAGAAGCTGTGGCGCAAGCTGAACCTGGCGCGAGCAGATCTACCCCGGCATCAAGGTGCGC	1497
DB	760	GlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyThrProGlyIleLysValArg	779
QY	1498	CAGCTGTGCAAGTGTGCGCGCGCCAGAGCCCTTACCGACATCTGTGCGCGCTGACCGAG	1557
DB	780	GlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGlu	799
QY	1558	GAGCGCGAGCTGAGCTGCGCGAGAACCGGAGATCTCTGCGCGAGCCCGTGCACCGGTG	1617
DB	800	GluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyVal	819
QY	1618	TACTACGACCCAGCAGGAGACCTGTGGCGCGAGATCCAGAGGAGGCGGACGACGAGTGG	1677
DB	820	TyTrpAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrp	839

QY 1678 ACCTACCAGATCTACACGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATG 1737  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 840 ThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMec 859  
 QY 1738 CGCACCGCCACACACACGCTGAAGACGCTGACCGAGCGCGTGCAGAGATCGCCATG 1797  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 860 ArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879  
 QY 1798 GAGACATCGTGATCTGGGGCAAGACCCCAAGTTCGGCTGCCATCCAGAGAGAC 1857  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 880 GluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGluThr 899  
 QY 1858 TGGGAGACCTGGTGGACCGACTACTCGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTG 1917  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 900 TrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919  
 QY 1918 AACACCCCGCCCTGTGAAGCTGTGTACAGCTGGAGAAGGAGGCCCATCATCGCGCGCG 1977  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 920 AsnThrProProLeuValLysLeuTyrTyrGlnLeuGluLysGluProIleValGlyAla 939  
 QY 1978 GAGACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTAC 2037  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 940 GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959  
 QY 2038 GTGACCGACCGGGCGCGAGAGATCGTCAGCGCTGACCGAGACCAACCAAGAGAC 2097  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 960 ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979  
 QY 2098 GAGCTGAGGCCATCCAGCTGGCGCTGCGAGGACGCGGCGAGCGAGTGAAACATCGTGACC 2157  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 980 AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999  
 QY 2158 GACAGCCAGTACCGCTGGGCATCATCCAGGCCCGCCGACCAAGAGCGAGCGAGCTG 2217  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1000 AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019  
 QY 2218 GTCAACACGATCATCGACGAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGGTGCC 2277  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1020 ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaIleProValPro 1039  
 QY 2278 GCCCACAAGGGCATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGC 2337  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1040 AlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059  
 QY 2338 AAGTGCTGTCTCGACGGCATCGAT 2364  
 Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 1060 LysValLeuPheLeuAspGlyIleAsp 1068

Search completed: June 2, 2005, 05:28:17  
 Job time : 228.167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 11:33:41 ; Search time 167.424 Seconds  
(without alignments)

Sequence: 1 gtcagccaccatgcccga.....gggctagcaccggtgaattc 2457  
11351.628 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 4555

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Deiop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh

-O=/cgn2\_1/USPFO\_spool/US09610313/runat\_31052005\_155136\_15139/app\_query.fasta\_1.7893

-DB=A\_Geneseq\_16Dec04 -QPMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0

-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi

-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15

-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000

-USPR=US09610313 @CGN 1 1 557/runat\_31052005\_155136\_15139 -NCPU=6 -ICPU=3

-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG

-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Geneseq16Dec04:\*

2: Geneseq1980s:\*

3: Geneseq1990s:\*

4: Geneseq2000s:\*

5: Geneseq2001s:\*

6: Geneseq2002s:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3723.5	81.7	1435	ADS19486	Ads19486 Modified
2	3701.5	81.3	3183	ADP84803	Adp84803 HIV-1 hyb
3	3668.5	80.5	1457	Adn36406	Adn36406 Human pro
4	3577	78.5	1003	AAB69289	Aab69289 HIV-1 non
5	3568	78.3	999	3 AAB69286	Aab69286 HIV-1 non
6	3563	78.2	1005	3 AAB69287	Aab69287 HIV-1 non
7	3527	77.4	998	6 AAE37601	Aae37601 HIV-1 sub
8	3520	77.3	3025	4 AAB86169	Aab86169 HIV-1 sub
9	3478	76.4	854	5 AAM48949	Aam48949 HIV-1 sub
10	3474.5	76.3	1000	3 AAB69282	Aab69282 HIV-1 non

11	3456.5	75.9	1002	2 AAW72993	Aaw72993 HIV isola
12	3454	75.8	1003	6 AAO30963	Aao30963 HIV pol p
13	3454	75.8	1003	8 ADN36413	Adn36413 HIV prote
14	3451.5	75.8	1002	1 Aap81861	Aap81861 Sequence
15	3451	75.8	1003	6 ABR55489	Abr55489 Amino aci
16	3449	75.7	1003	1 Aap60420	Aap60420 Sequence
17	3449	75.7	1003	1 Aap70861	Aap70861 Sequence
18	3443	75.6	1003	3 Aay70602	Aay70602 Codon opt
19	3443	75.6	1003	3 Aay70601	Aay70601 Corrected
20	3441	75.5	1003	5 AAO19387	Aao19387 Lymphaden
21	3435.5	75.4	1003	2 AAR08060	Aar08060 HIV-1 pol
22	3432	75.3	1015	1 Aap60347	Aap60347 HTLV-III
23	3432	75.3	1015	2 AAR43867	Aar43867 HTLV-III
24	3432	75.3	1015	4 AAB85993	Aab85993 Amino aci
25	3429	75.3	1003	3 AAB10047	Ab10047 HIV-1 pol
26	3429	75.3	1003	3 AAY70600	Aay70600 Wild type
27	3428	75.3	1016	2 AAR08062	Aar08062 AcNPV-HIV
28	3425	75.2	1015	2 AAR43875	Aar43875 HTLV-III
29	3421	75.1	1016	2 AAR08063	Aar08063 HIV-1 pol
30	3418	75.0	2033	2 AAR08056	Aar08056 HIV-1 pol
31	3414	75.0	1016	2 AAR08054	Aar08054 HIV-1 non
32	3412	74.9	1003	3 AAB69284	Aab69284 HIV-1 non
33	3411	74.9	2033	2 AAR08055	Aar08055 HIV-1 pol
34	3410	74.9	1003	8 ADP20076	Adp20076 Human imm
35	3406	74.8	739	5 AAU11874	Aau11874 HIV pol p
36	3406	74.8	1003	1 AAP61508	Aap61508 Sequence
37	3406	74.8	1003	2 AAR29705	Aar29705 pol gene
38	3406	74.8	1003	3 AAY77301	Aay77301 HIV-1 (AT
39	3406	74.8	1003	5 AAE35789	Aae35789 ARV-2 (9B
40	3406	74.8	1003	6 ABU63185	Abu63185 Pol prote
41	3406	74.8	1491	1 AAP91048	Aap91048 Transcrip
42	3403	74.7	1003	3 AAB69279	Aab69279 HIV-1 non
43	3401	74.7	1003	6 ABU57552	Abu57552 AIDS abso
44	3399	74.6	1012	6 ABU63325	Abu63325 Human lym
45	3396.5	74.6	1003	2 AAR08059	Aar08059 HIV-1 pol

ALIGNMENTS

RESULT 1

ADS19486  
ID ADS19486 standard; protein; 1435 AA.

AC ADS19486;

DT 18-NOV-2004 (first entry)

DE Modified HIV-1 GagPol protein.

Gene delivery; epithelial cell; respiratory system; mammal;  
lentiviral packaging system; expression vector; gagpol; gene expression;  
lentivirus; cystic fibrosis transmembrane conductance regulator; CFTR;  
cystic fibrosis; CF; deltaPD; trans epithelial potential difference;  
gene therapy; HIV-1; mutant; mutein.

Human immunodeficiency virus 1.

Synthetic.

US2004037780-A1.

26-FEB-2004.

23-AUG-2002; 2002US-00226638.

19-NOV-2001; 2001AU-00008942.

(PARS/) PARSONS D.

(ANSO/) ANSON D.

(LIMB/) LIMBERIS M.

(FULL/) FULLER M.

Parsons D, Anson D, Limberis M, Fuller M;

DR WPI: 2004-191004/18.  
 DR N-PSDB; ADS19485.  
 XX Delivering exogenous (e.g. CFTR) genes to epithelial cells in the  
 PT respiratory tract using a penetration agent and a lentiviral packaging  
 PT system, useful for treating e.g. cystic fibrosis.  
 XX  
 PS Disclosure; Fig 16; 45pp; English.  
 XX  
 CC The present invention relates to a method of delivering one or more  
 CC exogenous genes for expression in an epithelial cell in the respiratory  
 CC system of a mammal to give persistent expression of the gene in the  
 CC epithelial cell, using a lentiviral packaging (expression vector) system.  
 CC Also disclosed are (i) a recombinant lentiviral packaging system,  
 CC comprising a first nucleic acid molecule including a gag gene sequence  
 CC encoding a gag protein, and a second nucleic acid molecule including a  
 CC gagpol gene sequence encoding a GagPol protein (the gagpol gene sequence  
 CC has degenerative nucleotide changes in the frame shift sequence AUUUUU  
 CC to reduce the chance of a frame shift which switches expression of the  
 CC Gagpol protein to the gag protein in wild type lentivirus and the  
 CC packaging system additionally comprises a lentiviral vector nucleic acid  
 CC molecule which does not encode either the gag gene or the gagpol gene or  
 CC both, and (ii) a recombinant nucleic acid molecule encoding a lentiviral  
 CC gagpol gene having degenerative nucleotide changes in the frame shift  
 CC sequence AUUUUU to reduce the chance of a frame shift which switches  
 CC expression of the Gagpol protein to the gag protein in the wild type  
 CC lentivirus. The exogenous gene is expressed in sufficient numbers of  
 CC cells and amounts to provide an ameliorating effect for a respiratory  
 CC condition. The exogenous gene is cystic fibrosis transmembrane  
 CC conductance regulator (CFTR) and the condition is cystic fibrosis  
 CC ((CF) the expression of the CFTR gene is sufficient to provide a  
 CC significant shift of a reduced delta*apd* (transepithelial potential  
 CC difference) back to normal levels in the mammal). The cell is non-  
 CC terminally differentiated and is capable of differentiating into 2 or  
 CC more cell classes e.g. ciliated cells, non-ciliated cells, secretory  
 CC cells and basal cells. The exogenous gene may be enzymic. The gagpol gene  
 CC sequence has additionally degenerate nucleotide substitutions which  
 CC destabilise the hairpin structure associated with the frame shift event.  
 CC The gag gene is a truncation of the wild type gagpol gene so that it can  
 CC no longer be translated to form Gagpol. The lentivirus is HIV or HIV  
 CC derived. The method and recombinant lentiviral packaging system may be  
 CC used for delivering nucleic acids to epithelial cells in the respiratory  
 CC system of a mammal. In particular they may be used to deliver the CFTR  
 CC gene for the treatment of CF. It has been found that the delivery of a  
 CC recombinant lentivirus carrying an exogenous gene to the respiratory  
 CC system, following the delivery of a non-toxic amount of a penetrator  
 CC enhancer can provide persistence of expression of a gene product. The  
 CC present sequence represents modified HIV-1 GagPol protein.  
 XX  
 SQ Sequence 1435 AA;

Alignment Scores:  
 Pred. No.: 6,598-207 Length: 1435  
 Score: 3723.50 Matches: 698  
 Percent Similarity: 93.81% Conservative: 44  
 Best Local Similarity: 88.24% Mismatches: 40  
 Query Match: 81.75% Indels: 9  
 DB: 8 Gaps: 5

US-09-610-313B-32 (1-2457) x ADS19486 (1-1435)

QY 130 ATGAGGAGCTGCACCGAGCGCCAGCCCAACTTCTTCGCGAGGAGCTGGCGCTTCCCGCAG 249  
 DB MetLysAspCysThrGluArgGlnAlaAsnPheLeuArgGluAspLeuAlaPheProGln 442  
 QY 250 GGCAGGCGCCGCGAGTTCCCGCAGCGAGCAGAACCGCCCAACAGCCCCCAGCCGCGAG 309  
 DB GlyLysAlaArgLysPheSerSerGluGlnThrArgAlaAsnSerProileArgGlu 462  
 QY 310 CTGCGAGTG-----CGCGGCGACAAACCCCGCAGCGAGCGCGCGCGCGCGCGCGC 363  
 DB ArgGlnValTrpArgArgAspAsnAsnSerLeuSerGluAlaGlyAlaAspArgGlnGly 482  
 QY 364 ACCCTG-----AACTCTCCCGCAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAG 417  
 DB ThrValSerPheSerPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLys 502  
 QY 418 GTGGCGCGCCAGATCAAGAGGCGCTGTGTGACACCGCGCGCGCGAGCACCGTGTGGAG 477  
 DB IleGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGlu 522  
 QY 478 GAGATGACCTGCCCGCGCAAGTGGNAGCCCAAGATGATCGCGCGCATCGCGGCTTCATC 537  
 DB GluMetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIle 542  
 QY 538 AAGGTGCGCCAGTACGACACAGATCTGATCGAGATCTCGCGCAAGAGCCATCGGCACC 597  
 DB LysValArgGlnTyrAspGlnIleProIleGluLeuLysGlyHisLysAlaIleGlyThr 562  
 QY 598 GTGTGTGATCGCGCCCGCCCGTGAATCATCGCGCGCAACATGTCACCGCATCGGTGGC 657  
 DB ValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGly 582  
 QY 658 TGCACCTGAACTTCCCGCATCGACCCCATCGAGACCGTGCCTGCGAGCACTGAGAGCCGCG 717  
 DB CysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGly 602  
 QY 718 ATGAGCGCCCGCAAGGTGAAGTGGCGCTGCGCGAGGAGAGAGTCAAGGCGCTGACC 777  
 DB MetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuVal 622  
 QY 778 GCCATCTCGAGGAGATGAGAGGCGCGCAAGATCACCAAGATCGCGCGCGCGAGAACCCC 837  
 DB GluIleCysThrGluMetGluLysGluGlyLysSerLysIleGlyProGluAsnPro 642  
 QY 838 TACACACCCCGCTGTTCCCGCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897  
 DB TyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuVal 662  
 QY 898 GACTTCCGCGAGCTGAACAAAGCGCACCGACACTTCTGGAGGTGCAGCTGGGCATCCCC 957  
 DB AspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIlePro 682  
 QY 958 CACCCCGCGCGCTGAG 1017  
 DB HisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyr 702  
 QY 1018 TTCAGCGTCCCGCGATC 1077  
 DB PheSerValProLeuHisGluAspPheArgLysTyrThrAlaPheThrIleProSerIle 722  
 QY 1078 AACACAG 1137  
 DB AsnAsnGluThrProGlyThrArgTyrGlnTyrAsnValLeuProGlnGlnTrpLysGly 742  
 QY 1138 AGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCGCTTCCGCGCGCGCAAC 1197  
 DB SerProAlaIlePheGlnSerSerMetThrThrIleLeuGluProPheArgLysGlnAsn 762  
 QY 1198 CCGGAGATCGTGATCTACAG-----GCCCGCTGTAGCTGGCGAGCATCGTGAGATC 1251  
 DB ProAspLeuValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIle 782  
 QY 1252 GGCACGACCGCGCCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311





```
Db 1083 GlnLeuGluYsGluProIleGlyAlaGluThrPheYrValAspGlyAlaAsn 1102
Qy 2002 CGCAGACCAAGATCGGCAAGCGCGGTACGTGACCGACCGCGCGCGCAAGATCGTG 2061
Db 1103 ArgGluThrYsLeuGlyYsAlaGlyYrValThrAspArgGlyArgGlnYsValVal 1122
Qy 2062 AGCTGACCGAGACCAACCAAGACCGAGCTGCGAGGCCATCCAGCTGGCCCTGCAG 2121
Db 1123 ProLeuThrAspThrThrAsnGlnYsThrGluLeuGlnAlaIleHisLeuAlaLeuGln 1142
Qy 2122 GACAGCGCAGCGAGGTGACATCGTACCGACAGCCAGTACGCGCTGGCGATCATCCAG 2181
Db 1143 AspSerGlyLeuGluValAsnIleValThrAspSerGlnYrAlaLeuGlyIleIleGln 1162
Qy 2182 GCCCAGCGCAGACAGAGCGAGCGAGCTGGTGAACAGATCATCGAGCGATCATCAAG 2241
Db 1163 AlaGlnProAspYsSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIleYs 1182
Qy 2242 AAGGAGAGGTGTACCTGAGCTGGTGGTCCCGCCCAAGGGCATCGCGCGCAAGCAGCAG 2301
Db 1183 YsGlnYsValYrLeuAlaTriValProAlaHisYsGlyIleGlyGlyAsnGluGln 1202
Qy 2302 ATCGACAGCTGTGAGCAGGCGCATCGCGAAGTGTCTTCTCGAGCGGATCGAT 2358
Db 1203 ValAspYsLeuValSerAlaGlyIleArgYsValLeuPheLeuAspGlyIleAsp 1221

RESULT 3
ADN36406
ID ADN36406 standard; protein; 1457 AA.
XX AC ADN36406;
XX DT 15-JUL-2004 (first entry)
XX DE Human protein for anti-HIV vaccine.
XX KW anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX OS Homo sapiens.
XX PN W0204035006-A2.
XX PD 29-APR-2004.
XX PF 17-OCT-2003; 2003WO-US033112.
XX PR 18-OCT-2002; 2002US-0419465P.
XX PA (AARO-) AARON DIAMOND AIDS RES CENT.
XX PI Huang Y, Ho DD, Chen Z;
XX DR WPI; 2004-348328/32.
XX DR N-PSDB; ADN36405.
XX CC Nucleic acid vector comprising at least one HIV sequence operably linked
PT to a promoter and encoding a protein that does not assemble into viral
PT particles, useful in immunizing a subject against HIV infection.
XX PS Disclosure; SEQ ID NO 20; 166pp; English.
XX CC The invention relates to a nucleic acid vector comprising at least one
CC HIV sequence operably linked to a promoter and encoding a protein that
CC does not assemble into viral particles. The nucleic acid vector is useful
CC in immunizing a subject against HIV infection. This sequence corresponds
CC to a peptide used in the invention.
XX SQ Sequence 1457 AA;

Alignment Scores: 1e-203 Length: 1457
Pred. No.: 3668.50 Matches: 702
Score:
```

```
Percent Similarity: 91.36% Conservative: 28
Best Local Similarity: 87.86% Mismatches: 47
Query Match: 80.54% Indels: 23
DB: 8 Gaps: 5
US-09-610-313B-32 (1-2457) x ADN36406 (1-1457)
Qy 13 ATGCGCCGAGCCATGAGCCAGGCCAGCCCAACATCTGTATGCGCGCAGCAACTTC 72
Db 383 LeuAlaGluAlaMetSerGlnAla---AsnGlyThrIleLeuMetGlnArgSerAsnPhe 401
Qy 73 AAGGCCCCCAAGCGCATCATCAAGTCTTCACTGCGCGCAAGGAGGGCCACATGCCCGC 132
Db 402 YsGlySerYsArgIleValYsCyPheAsnCySgLYsGluGlyHisIleAlaArg 421
Qy 133 AACTGCGCGCCCGCCCAAGAGAGGGTGTCTGGAAGTGGCGCAAGGAGGGCCACAGATG 192
Db 422 AsnCysArgAlaProArgYsYsGlyCysTrpYsCysGlyYsGluGlyHisGlnMet 441
Qy 193 AAGGACTGCACCGAGCGCCAGGCCCACTTCTTCGCGAGGACCTTCCCCCAGGGC 252
Db 442 YsAspCysThrGluArgGlnAlaAsnPheLeu-GlyYsIleTrpProProHisYsG1 461
Qy 253 AAGGCGCGGAGTTCGCCAGCGAGCAGAACCGCGCCCAACAGCCCGCAGCGCGAGTG 312
Db 461 YArgProGlyYsAsnPheLeuGlnSerArgProGluProThrAlaProAlaGluSerPh 481
Qy 313 CAGGTGCGCGCGCAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 481 eGlyPheGluThrThrProAlaProYsGlnGluProYsAspArgGluProLeuTh 501
Qy 373 TTC-----CCCCAGATCACCCCTGTGGCA 395
Db 501 rSerLeuYsSerLeuPheGlySerAspProLeuSerGlnProGlnIleThrLeuTrpG1 521
Qy 396 GCGCGCCCTGTGAGCATCAAGTGGCGCGCGAGATCAAGAGGCCCTCTCTGGACACCGG 455
Db 521 nArgProLeuValSerIleArgValGlyGlyGlnIleYsGluAlaLeu----- 538
Qy 456 CGCGCAGCACACCGTCTGCGAGGAGATGAGCTGCGCGCAAGTGGAGGCCAAGATGAT 515
Db 539 ----AspAspThrValLeuGluGluValAsnLeuProGlyYsTrpYsProYsMecI1 557
Qy 516 CGCGCGCATCGCGCGCTTCATCAAGTGCGCCAGTACGACCATCTGTATGCGAGATCTG 575
Db 557 eGlyGlyIleGlyPheIleYsValArgGlnTrpAspGlnIleProIleGluIleCy 577
Qy 576 CGCGAAGAGCCCATCGGCACCGTCTGTGATCGCGCCCGCCCGCGCGCGCGCGCGCG 635
Db 577 sGlyYsYsAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyAr 597
Qy 636 CAACATGCTGACCCAGCTGGCTGCACCTGAACTTCCCATCGACCGCCCATCGAGACCGT 695
Db 597 gAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrI1 617
Qy 696 GCCCGTGAAGCTGAAGCCCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 755
Db 617 eProValYsLeuYsProGlyMetAspGlyProArgValYsGlnTrpProLeuThrG1 637
Qy 756 GGAGAGATCAAGCCCTGACCGCGCATCTCGAGAGAGATGGAGAGGAGGAGGAGATCAC 815
Db 637 uGluYsIleYsAlaLeuThrAlaIleCysAspGluMetGluYsGluYsIleTh 657
Qy 816 CAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 875
Db 657 rYsIleGlyProGluAsnProYrAsnThrProValPheAlaIleYsYsYsAspSe 677
Qy 876 CACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTGAACAAGCGCGCGCGCGCGCTCTG 935
Db 677 rThrYsTrpArgYsLeuValAspPheArgGluLeuAsnYsArgThrGlnAspPheTr 697
Qy 936 GGAGGTGACGTGGGCATCCCCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 995
```



Db 697 pGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrVa 717  
Qy 996 GCTGACGTGGCGAGCGCTACTTTCAGCGTGGCCCTGGACGAGGACTTCCGCAATACAC 1055  
Db 717 lIleuAspValGlyAspAlaTyPheSerValProLeuTyrGluAspPheArgLysTyThr 737  
Qy 1056 GCCTTTACCATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTACCGTACAAGCT 1115  
Db 737 rAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnVa 757  
Qy 1116 GCTGCCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCCT 1175  
Db 757 lLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnCysSerMetAlaLysIleLe 777  
Qy 1176 SGAGCCCTTCGCGCGCGCAACCCGAGATCGTGATCTACAG-----GCCCCCTGTA 1229  
Db 777 uGluProPheArgAlaGlnAsnProGluIleValIleTyrGlnTyrGlyAspAspLeuTy 797  
Qy 1230 CGTGGCGAGCGACTGGAGATCGGCAGCACCGCCCAAGATCGAGGAGCTGCGCAAGCA 1289  
Db 797 rValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHI 817  
Qy 1290 CTGCTGGCTGGGCTTTCACACCCCGCACAAAGACCAAGAGAGAGCCCTTCTCT 1349  
Db 817 sLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLe 837  
Qy 1350 G-----CCCATCGAGCTGCACCCCGACAAAGTGGACCGTGCAGCCATCGAGCTGCCGA 1403  
Db 837 uTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGI 857  
Qy 1404 GAAGGAGAGCTGACCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAACCTGGCGCAG 1463  
Db 857 uLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSe 877  
Qy 1464 CCAGATCTACCCCGCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCGCGCAAGGC 1523  
Db 877 rGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAl 897  
Qy 1524 CTTGACCGACATCGTGGCCCTGACCGAGAGCGCGAGCTGGAGCTGGCGGAGAACCGGA 1583  
Db 897 aLeuThrAspIleIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGI 917  
Qy 1584 GATCTCGCGAGCGCGTGCAGCGGTGTACTACGACCCCGCAAGGACCTGGTGGCGCA 1643  
Db 917 uLleLeuLysGluProValHisGlyAlaTyrTyrAspProSerLysAspLeuIleAlaGI 937  
Qy 1644 GATCAGAGCAGGCGCCACGACCTGACCTACAGATCTACAGAGGAGCCCTTCAAGAA 1703  
Db 937 uIleGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAs 957  
Qy 1704 CTTGAAGACCGCAAGTACGCCAAGATGGCGACCCGCCACACACGACGAGTGAAGCAGCT 1763  
Db 957 nLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLe 977  
Qy 1764 GACCGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTGATCTGGGCAAGACCCCA 1823  
Db 977 uThrGluAlaValGlnLysIleSerMetGluSerIleValIleTrpGlyLysIleProTy 997  
Qy 1824 GTTCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGCACTACTGGCAGGC 1883  
Db 997 sPheArgLeuProIleProLysGluThrTrpGluThrArgTrpThrAlaTyrTrpGlnAl 1017  
Qy 1884 CACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTGTACA 1943  
Db 1017 aThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGI 1037  
Qy 1944 GCTGAGAGAGCGCCATCATCGCGCGGAGACCTTCTAGTGGAGCGCGCCCAACCG 2003  
Db 1037 nLeuGluLysAspProIleAlaGlyValGluThrPheTyrValAspGlyAlaAsnAr 1057  
Qy 2004 CGAGACCAAGATCGCGAGCGCGCTAGCTGACCGACCGCGCGCGCAGAGATCGTAG 2063  
Db 1057 gGluThrLysMetGlyLysAlaGlyTyValThrAspArgGlyArgGlnLysIleValSe 1077

Qy 2064 CCTGACGAGACCCACCAACGAGAGCCGAGCTGCAGGCCATCCAGCTGGCCCTCAGGA 2123  
Db 1077 rLeuThrGluThrThrAsnGlnLysThrGluGlnAlaIleCysLeuAlaLeuGlnAs 1097  
Qy 2124 CAGCGGACGAGGTGAACATCGTGACCGACGACGACGACGACGACGACGACGACGACG 2183  
Db 1097 pSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAl 1117  
Qy 2184 CCAGCCCGACAGGACGAGCGAGCTGGTGAACACGATCATTCGAGCAGCTGATCAAGAA 2243  
Db 1117 aGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuLysLy 1137  
Qy 2244 GGAGAGGTGTACTGAGCTGGTGGCCCGCCCAAGGGCATCGCGGCAAGCAGCAGAT 2303  
Db 1137 sGluArgValTyrLeuSerTrpValProAlaHisGlyIleGlyGlyAsnGluGlnVa 1157  
Qy 2304 CGAAGCTGTGAGCAAGGCGCATCCGCAAGGTGTCTCTGACGCGCATCGAT 2358  
Db 1157 lAspLysLeuValSerAsnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 1175

RESULT 4  
AAB69289  
ID AAB69289 standard; protein; 1003 AA.  
XX AAB69289;  
XX 12-SEP-2003 (revised)  
DT 20-APR-2001 (first entry)  
XX HIV-1 non-subtype B clone 94IN476-104 pol protein.  
XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
KW vif; vpr; tat; rev; nef; vaccine.  
XX Human immunodeficiency virus 1.  
XX WO200026416-A1.  
XX 11-MAY-2000.  
XX 25-OCT-1999; 99WO-US024837.  
XX 02-NOV-1998; 98US-00184418.  
XX (UABR-) UAB RES FOUND.  
XX Hahn BH, Shaw GM, Gao F;  
XX WPI; 2000-365651/31.  
XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
PT type 1 useful for detecting and treating AIDS comprises a specific  
PT nucleotide sequence.  
XX Claim 41; Fig 15; 131pp; English.  
XX The present invention provides the protein and coding sequences for a  
CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
CC isolates. The sequences shown include the near full-length coding  
CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
CC rev and nef proteins. These can be used to detect the presence of HIV-1  
CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
CC These antibodies can be used in vaccines to prevent and treat HIV  
CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
XX SQ Sequence 1003 AA;  
Alignment Scores:  
Pred. No.: 1.8e-198 Length: 1003  
Score: 3577.00 Matches: 673  
Percent Similarity: 97.49% Conservative: 26  
Best Local Similarity: 93.86% Mismatches: 14

Query Match: 78.53% Indels: 4  
DB: 3 Gaps: 2  
US-09-610-313B-32 (1-2457) x AAB69289 (1-1003)

220 TTCTTCCGCGAGGACCTGCGCTTCCCGCAGGCAAGCGCGGAGTTCCTCCCGAGCGCAG 279  
Db 1 PhePheArgGluAenLeuAlaPheProGlnGlyGluAlaArgGluPheProSerLysGln 20  
280 AACCGCGCAACAGCCCCCAGCGCGCGAGCTGCGAGTGGCGCGCGCAACACCCCGCAGC 339  
Db 21 AlaargAlaenSerProThrSerArgGluLeuGlnValGlnGlyAaspAsnProArgSer 40  
340 GAGCGCGCGCGCGAGCGCGCAGGCGACCTGAACTTCCCGCAGATCACCTGTGCGAGCGC 399  
Db 41 GluAlaGlyValGluArgGlnGlyThrLeuAenPheProGlnIleThrLeuTrpGlnArg 60  
400 CCCCTGTGTGACATCAGTGGCGCGCGCGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATCGGC 519  
Db 61 ProleuValSerIleLysValGlyGlnIleLysGluAlaLeuLeuAenPheThrGlyAla 80  
460 GACGACACCGTGTGCGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATCGGC 519  
Db 81 AspAspThrValLeuGluGluIleAlaLeuProGlyArgTrpLysProLysMetIleGly 100  
520 GGCATCGCGCGCTTCATCAAGGTGCGCGCAGTACGACCAAGATCTCGATCGAGATCTGCGGC 579  
Db 101 GlyIleGlyGlyPheIleLysValArgGlnIleThrAspGlnIleLeuIleGluIleCysGly 120  
580 AAGAAGCGCATCGCGCGCGCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 639  
Db 121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAenIleIleGlyArgAasp 140  
640 ATGTGTGACCGAGTGGCTGACCTGAACTTCCCATCTCCAGCCCGCATCGAGACCGTGC 699  
Db 141 MetLeuThrGlnLeuGlyCysThrLeuAenPheProIleSerProIleGluThrValPro 160  
700 GTGAAGCTGAAGCG 759  
Db 161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180  
760 AAGATCAAGCG 819  
Db 181 LysIleLysAlaLeuThrGluIleCysLysGluMetGluLysGluGlyLysIleThrLys 200  
820 ATCGCGCGCGAGAACCTTCAACACACCGCGTGTTCGCCATCAAGAGAGAGAGAGAGAG 879  
Db 201 IleGlyProGluAenProTyAsnThrProValPheAlaIleLysArgLysAspSerThr 220  
880 AAGTGGCGCAAGCTGTGGACTTCGCGAGCTGAACAGCGCACCCAGGACTTCTTGGGAG 939  
Db 221 LysTrpArgLysLeuValAspPheArgGluLeuAenLysArgThrGlnAaspPheTrpGlu 240  
940 GTGAGCTGGCGATCCCG 999  
Db 241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeu 260  
1000 GACGTGGCGCGCGCTACTTCCAGCGTCCCGTGGACGAGGACTTCCGCAAGTACACCGCC 1059  
Db 261 AspValGlyAspAlaTyPheSerValProLeuAaspGluGlyPheGlyLysTyThrAla 280  
1060 TTCAACATCCCGCAGCATCAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119  
Db 281 PheThrIleProSerIleAenAenGluThrProGlyIleArgTyGlnTyAsnValLeu 300  
1120 CCCGAGCGCTGAAGGCG 1179  
Db 301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320  
1180 CCCTTCCG 1233  
Db 321 PropheArgAlaArgAenProLysIleValIleTyGlnTyMetAspLeuTyVal 340

1234 GGCAGCGACCTGGAGATCGGCCACGACCGCCCAAGATCGAGGAGCTGCGCAAGCACCTG 1293  
Db 341 GlySerAspLeuGluIleGlyHisArgAlaLysIleGluLeuArgAlaHisLeu 360  
1294 CTGCGCTGGGCTTCCACCCCGCAGCAAGAGACACAGAGAGGAGCCCGCTTCCCTG--- 1350  
Db 361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380  
1351 ---CCCATCGAGCTGCGCCCGCAGCAAGTGAAGCTGCGCGCGCGCGCGCGCGCGCG 1407  
Db 381 MetGlyTyGluLeuHisProAspLysTrpThrValGlnProIleLysLeuProGluLys 400  
1408 GAGAGCTGGACCGTGAACAGCATCCAGAAGCTGTGGGCAAGCTGAACCTGGGCGCAGC 1467  
Db 401 AspSerTrpThrValAenAspIleGlnLysLeuValGlyLysLeuAenTrpAlaSerGln 420  
1468 ATCTACCCCGCATCAAGTGGCGCGCGAGCTGCGAGCTGCGCGCGCGCGCGCGCGCG 1527  
Db 421 IleTyProGlyIleLysValArgGlnLeuCysLysLeuLeuArgLysAlaLysAlaLeu 440  
1528 ACCGACATCTGCGCCCTGACCGAGGCGCGCGAGCTGCGCGCGCGCGCGCGCGCGCG 1587  
Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAenArgGluIle 460  
1588 CTGCGCGAGCGCGTGCACGCGCTGTACTAGACCCCGCAGCAAGAGCCTGTGTGGCGAG 1647  
Db 461 LeuLysGluProValHisGlyValTyTyAspProSerLysAspLeuIleAlaGluIle 480  
1648 CAGAGCAGCG 1707  
Db 481 GlnLysGlnGlyHisAspGlnTrpThrTyGlnIleTyGlnGluProPheLysAenLeu 500  
1708 AAGACCGCGCAAGTACGCGAGATCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1767  
Db 501 LysThrGlyLysTyAlaLysMetArgThrAlaHisThrAenAspValLysGlnLeuThr 520  
1768 GAGCGCTGCAAGAGATCGCCATGAGGAGCATCTGTGATCTGGGGCAAGACCCCGAGTTC 1827  
Db 521 GluAlaValGlnLysIleAlaIleGluSerIleValIleTrp\*\*LysThrProLysPhe 540  
1828 CGCTGCG 1887  
Db 541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpThrTrpThrAspTyTrpGlnAlaThr 560  
1888 TGGATCCCGCGAGTGGAGTTCGTGAACACCGCGCGCGCGCGCGCGCGCGCGCGCG 1947  
Db 561 TrpIleProAspTrpGluPheValAsnThrProLeuValLysLeuTrpTyGlnLeu 580  
1948 GAGAAGAGCGCGCATCATCGCGCGCGAGACCTTCTACGTGGACGCGCGCGCGCGCG 2007  
Db 581 GluLysGluProIleValGlyAlaGluThrPheTyValAspGlyAlaAaenArgGlu 600  
2008 ACCAAGATCG 2067  
Db 601 ThrLysValGlyLysAlaGlyTyValThrAspArgGlyArgGlnLysIleValSerLeu 620  
2068 ACCGAGACCCCAACAGAGAGCGAGCTGCGAGCGCATCCAGCTGGCGCGCGCGCGCG 2127  
Db 621 ThrGluThrThrAenGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSer 640  
2128 GGCAGCGAGGTGAACATCTGTGACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 2187  
Db 641 GlyThrGluValAenIleValThrAspSerGlnTyAlaLeuGlyIleIleGlnAlaGln 660  
2188 CCGGAGAGAGCG 2247  
Db 661 ProAspLysSerGluSerGluLeuValAenGlnIleIleGluGlnLeuIleAenLysGlu 680  
2248 AAGTGTACCTGAGCTGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2307  
Db 681 ArgValTyLeuSerTrpValProAlaHisLysGlyIleGlyLysGlnGluValAsp 700  
2308 AAGCTGTGAGCAAGGCGCATCCCGAAGGTGCTGTCTCTGCGCGCGCGCGCGCGAT 2358

Db 701 ArgLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleasp 717  
 :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: :::::::::::::: ::::::::::::::

RESULT 5  
 ID AAB69286  
 XX AAB69286 standard; protein; 999 AA.  
 XX AAB69286;  
 XX AC  
 XX 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 XX HIV-1 non-subtype B clone 96ZM651-8 pol protein.  
 XX  
 XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 XX  
 XX Human immunodeficiency virus 1.  
 XX  
 XX WO200026416-A1.  
 XX  
 XX 11-MAY-2000.  
 XX  
 XX 25-OCT-1999; 99WO-US024837.  
 XX  
 XX 02-NOV-1998; 98US-00184418.  
 XX  
 XX (UABR-) UAB RES FOUND.  
 XX  
 XX Hahn BH, Shaw GM, Gao F;  
 XX  
 XX WPI; 2000-365651/31.  
 XX  
 XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.  
 XX  
 XX Claim 41; Fig 15; 131pp; English.  
 XX  
 XX The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)  
 XX  
 XX Sequence 999 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 5,976-198 Length: 999  
 Score: 3568.00 Matches: 673  
 Percent Similarity: 97.21% Conservative: 24  
 Best Local Similarity: 93.86% Mismatches: 16  
 Query Match: 78.33% Indels: 4  
 Db: 3 Gaps: 2  
 US-09-610-313B-32 (1-2457) x AAB69286 (1-999)  
 QY 220 TTCTTCCGCGAGACTGGCCCTTCCCGAGGCAAGCCGCGAGTCCCGAGCGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20  
 QY 280 AACCGCGCCACAGCCCAACAGCCGCGAGTGCAGGTGCGCGGCAACACCCCGCAGC 339  
 Db 21 AlaArgAlaAsnSerProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSer 40  
 QY 340 GAGGCGGCGCGAGCGCCAGCGGACCTGAACTTCCCGCAGATCACCTGTGGCAGCGC 399  
 Db 41 GluAlaGlyValGluArgGlnGlySerLeuAsnPheProGlnIleThrLeuTrpGlnArg 60  
 QY 400 CCCTGGTGAGCATCAAGGTGGCGGCCAGATCAAGGAGGCCCTGTGTGACACCGCGCC 459

Db 61 ProLeuValSerIleLysValGlyGlnIleLysGluAlaLeuLeuAspThrGlyAla 80  
 QY 460 GACGACACCGTCTGGAGGAGATGAGCTGCCCGCAAGTGGAAGCCCAAGATGATCGGC 519  
 Db 81 GlyAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100  
 QY 520 GGCATCGGGCGCTTCATCAAGGTGGCGGACGACGACGAGATCCTCATCGAGATCTGCGGC 579  
 Db 101 GlyIleGlyGlyPheIleGluValArgGlnTrpAspGlnIleProMetGluIleCysGly 120  
 QY 580 AAGAAGCGCATCGGCACCGTGTGATCGGCGCCCGCCACCGCGTGAACATCATCGCGCGCAAC 639  
 Db 121 LysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsn 140  
 QY 640 ATGTGACCCAGCTGGGTGACCTGAACCTCCCATCCCATCAGCCCATCGAGACCGTGC 699  
 Db 141 MetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValPro 160  
 QY 700 GTGAGCTGAAGCCCGCATGAGCGCCCGCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 759  
 Db 161 ValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGlu 180  
 QY 760 AAGATCAAGCGCTGACCGCCATCTCGAGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 819  
 Db 181 LysIleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLys 200  
 QY 820 ATCGGCGCGGAGAACCTTACAACACCGCTGTTGCGCATCAAGAGAGAGAGAGAGAGAGAG 879  
 Db 201 IleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLysLysAspSerThr 220  
 QY 880 AAGTGGCGCAAGCTGGTGGACTTCCCGAGCTGAACAGAGCCAGCCAGAGCTTCTGGGAG 939  
 Db 221 LysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGlu 240  
 QY 940 GTGACAGCTGGGCATCCCCACCGCCCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 999  
 Db 241 ValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeu 260  
 QY 1000 GACGTGGCGGAGCCCTACTTACGTGCTGCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059  
 Db 261 AspValGlyAspAlaTyPheSerValProLeuAspGluSerPheArgLysTrpThrAla 280  
 QY 1060 TTCACATCCCGAGCATCAACAG 1119  
 Db 281 PheThrIleProSerThrAsnAsnGluThrProGlyIleArgTrpGlnTrpAsnValLeu 300  
 QY 1120 CCCCAGGCGTGAAGGCGAGCCCGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1179  
 Db 301 ProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGlu 320  
 QY 1180 CCCTTCCCGCGCGCAACCGAGATCGTATCTACAG-----GCCCGCTGTACGTG 1233  
 Db 321 ProPheArgAlaGlnAsnProAspIleValIleTyGlnTrpMetAspLeuTyVal 340  
 QY 1234 GCGAGCGACCTGGAGATCGGCAGCAGCCGCGAGAGATCGAGGAGCTGCGAGAGAGAGAGAG 1293  
 Db 341 GlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGlnHisLeu 360  
 QY 1294 GTGCGCTGGGGTTCACACCGCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1350  
 Db 361 LeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrp 380  
 QY 1351 ---CCCATCGAGCTGCACCCGCAAGTGGAGCGCTGAGCGCCCATCGAGCTGCCCGAGAGAG 1407  
 Db 381 MetGlyTyTrpGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuAlaGluLys 400  
 QY 1408 GAGAGCTGAGCGTGAACGAGATCCAGAGAGCTGGGGCAAGCTGAACCTGGCGGAGAGAGAG 1467  
 Db 401 AspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGln 420  
 QY 1468 ATCTACCCCGGCATCAAGGTGGCGGCGCATGTGTGAAGAGCTGTGCGCGGCGCCAGAGCCCTG 1527

Db 421 IleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeu 440  
 Qy 1528 ACCGACATCGTGGCCCTGACGAGAGAGCCGAGCTGGAGCTGGCGAGAACCGGAGATC 1587  
 Db 441 ThrAspIleValProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnLysGluIle 460  
 Qy 1588 CTGGCGAGCCGCTGACGGGTGTACTACGACCCGAGCAAGAGCTGGTGGCGAGATC 1647  
 Db 461 LeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIle 480  
 Qy 1648 CAGAAGCAGGCGCCAGCAGCTGACCTACAGAGCTTACAGAGGAGCCCTTCAAGAACCTG 1707  
 Db 481 GlnLysGlnGlyHisAspGlnTrpTyrGlnIleTyrGlnGlnProPheLysAsnLeu 500  
 Qy 1708 AAGACCGCAAGTACGCCAAGATCGCACCCGACACCAACGACGCTGAAGCAGCTGACC 1767  
 Db 501 LysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThr 520  
 Qy 1768 GAGCCGTGCGAGAGATGCCATCGAGAGCATCTGTGATCTGGGGCAGACCCCAAGTTC 1827  
 Db 521 GluAlaValGlnLysIleAlaLeuGluSerIleValIleTyrGlyLysIleProLysPhe 540  
 Qy 1828 CGCTGTCCTCCACAGAGAGCTGGAGACCTGTGGACCGACTACTGGCAGGCCAC 1887  
 Db 541 ArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpTrpTrpTrpTrpTrpTrp 560  
 Qy 1888 TGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTGTGTACAGCTG 1947  
 Db 561 TrpIleProGluTrpGluPheValAsnThrProLeuLeuValLysLeuTrpTyrGlnLeu 580  
 Qy 1948 GAGAAGAGCCCATCATCTGCGCGCGAGACCTTCTACGTGAGCGCGCGCCCAACCGGAG 2007  
 Db 581 GluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGlu 600  
 Qy 2008 ACCAAGATCGCAGCGCGGTAGTGCACCGAGCTGGACCGCGCGGCGAGATCGTGGCTG 2067  
 Db 601 ThrLysLeuGlyLysAlaGlyTyrIleThrAspArgGlyArgGlnLysIleValThrLeu 620  
 Qy 2068 ACCGAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGC 2127  
 Db 621 ThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSer 640  
 Qy 2128 GGCAGCAGGTGAATCGTACCGACAGCAGCGAGTACGCGCTGGCGCATCATCCAGGCCAG 2187  
 Db 641 GlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaHis 660  
 Qy 2188 CCCACAGACGAGAGCGAGCTGGTGAACAGATCATCGAGCAGCTGATCAAGAGGAG 2247  
 Db 661 ProAspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuLysLysGlu 680  
 Qy 2248 AAGGTGTACCTGAGCTGGTGGCCGCCACCAAGGCGCATCGCGGCAACGAGCAGATCGAC 2307  
 Db 681 ArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysAsnGluGlnValAsp 700  
 Qy 2308 AAGTGTGTGACCAAGGCGCATCCGCAAGTGTGTCTTCTGGAGCGCATCGAT 2358  
 Db 701 LysLeuValSerLysGlyIleArgLysValLeuPheLeuAspGlyIleAsp 717

RESULT 6

AAB69287  
 ID AAB69287 standard; protein; 1005 AA.  
 XX  
 AC AAB69287;  
 XX  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 XX HIV-1 non-subtype B clone 96ZM751-3 pol protein.  
 XX HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; net; vaccine.  
 XX  
 XX Human immunodeficiency virus 1.  
 OS

XX WO200026416-A1.  
 FN 11-MAY-2000.  
 PD 25-OCT-1999; 99WO-US024837.  
 XX 02-NOV-1998; 98US-00184418.  
 XX (UABR-) UAB RES FOUND.  
 PA Hahn BH, Shaw GM, Gao F;  
 PI WPI; 2000-365651/31.  
 DR Novel genomic nucleic acids of non-subtype B human immunodeficiency virus  
 PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.  
 XX Claim 41; Fig 15; 131pp; English.  
 PS  
 XX The present invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection.. (Updated on 12-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 1005 AA;  
 Alignment Scores:  
 Pred. No.: 1,16e-197 Length: 1005  
 Score: 3563.00 Matches: 677  
 Percent Similarity: 96.13% Conservative: 18  
 Best Local Similarity: 93.64% Mismatches: 18  
 Query Match: 78.22% Indels: 10  
 DB: 3 Gaps: 3  
 US-09-610-313B-32 (1-2457) x AAB69287 (1-1005)

Qy 220 TTCTTCCGAGGACCTGGCTTCCCGGAGCAAGCCCGGAGTTCCTCCAGCGAGCAG 279  
 Db 1 PhePheArgGluAsnLeuAlaPheProGluGlyGluAlaGlyGluLeuProSerGluGln 20  
 Qy 280 AACCGCGCC-----AACAGCCCGCCAGCCGCGAGCTGCAGTGCAGTGGCC 321  
 Db 21 ThrArgAlaAsnSerProThrSerSerAsnSerProThrSerArgGluLeuGlnValArg 40  
 Qy 322 GCGGACAAACCCCGCAGCGAGCGCGCGAGCGCGGAGCGCGGAGCGCGGAGCGCGGAG 381  
 Db 41 GlyAspAsnProCysProGluAlaGlyAlaGluArgGlnGlyThrLeuAsnCysProGln 60  
 Qy 382 ATCAACCTGTGGCAGCGCCCTTGTGTGAGCATCAAGGTGGCGCGGAGCGCGGAGCGCGC 441  
 Db 61 IleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnLysGluAla 80  
 Qy 442 CTGCTGGACACCGCGCGCGAGCAGCACCGTGTGGAGAGATGAGCTGCCCGGCAAGTGG 501  
 Db 81 LeuLeuAspThrGlyAlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrp 100  
 Qy 502 AAGCCCAAGATGATCGCGCGCATCGCGCGCTTTCATCAAGGTGGCGCGGAGCGCGGAG 561  
 Db 101 LysProLysMetIleGlyGlyIleGlyPheLysValArgGlnTyrAspGlnIle 120  
 Qy 562 CTGATCGAGATCTCGCGCAAGAAGCGCATCGGACCGCTGTGTGATCGGCGCGGAGCGCGC 621  
 Db 121 LeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProVal 140  
 Qy 622 AACATCATCGCGCGCAGACATGCTGACCGAGTGGGTGCGACCTGAACTTCCCATCAGC 681  
 Db 141 AsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPheProIleSer 160

QY	682	CCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGCATGCAGCGCCCAAGGTGAAGCAG	741
DB	161	ProIleGluThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGln	180
QY	742	TGGCCCTGACCGAGGAGAAAGATCAAGGCCTGACCGCCATCTGCGAGGAGATGGAGAAG	801
DB	181	TrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGluMetGluLys	200
QY	802	GAGGGCAAGATCACCAAGATCGGCCCGAGAAACCCCTACAACACCCCGTTCGCCCATC	861
DB	201	GluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIle	220
QY	862	AAGAAGAAGACAGCACCAAGTGCAGCAAGCTGGTGACTTCCCGCAGCTGGAACAAGCGC	921
DB	221	LysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArg	240
QY	922	ACCAGGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCCTCGAAGAAGAAG	981
DB	241	ThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLys	260
QY	982	AAGAGCTGACCGTCTGGAGCTGGGGCGACGCCTACTTCAGCGTGCCTTGGACGAGGAC	1041
DB	261	LysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluGly	280
QY	1042	TTCGGCAAGTACACCGCTTCCACCATCCCGAGCATCAACAACGAGACCCCGCGCATCCGC	1101
DB	281	PheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArg	300
QY	1102	TACAGTACAACTGTGTCGCCCGAGGTGGAAGGCGAGCCCGACATCTTCCAGAGCAGC	1161
DB	301	TyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePheGlnSerSer	320
QY	1162	ATGACCAAGATCTGGAGCCCTTCCGCGCCGCGACCCCGAGATCGTGATCTACCAG---	1218
DB	321	MetIleLysIleLeuGluProPheArgThrGlnAsnProGluIleValIleTyrGlnTyr	340
QY	1219	--GCCCCCTGTACGTGGGCGAGCCTCGAGATCGGCAGCACCGCGCAAGATCGCAG	1275
DB	341	MetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlu	360
QY	1276	GAGCTGGCGAAGCACCTGTGCTGCGTGGGGCTTCAACCCCGCACAGAAGCACCAGAAG	1335
DB	361	GluLeuArgGluHisLeuLeuAsgTrpGlyPheThrThrProAspLysLysHisGlnLys	380
QY	1336	GAGCCCGCTTCCTG-----CCCATCGAGTGCACCCCGACAAGTGGACCGTGCAGCCC	1389
DB	381	GluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTyrThrValGlnPro	400
QY	1390	ATCGAGTGGCCGAGAGAGAGCTGCACCGTGAACGACATCCAGAGCTGGTGGGCAAG	1449
DB	401	IleLysLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLys	420
QY	1450	CTGAACCTGGCGACCGACAGATCTACCCCGGCATCAAGGTGGCCGACGTGTGCAAGCTG	1509
DB	421	LeuAsnTrp***SerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeu	440
QY	1510	CGCGGCGCCAAGGCCCTGACCGAATCGTGCCTCTGACCGAGAGGCGCGAGCTGGAGCTG	1569
DB	441	ArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLeu	460
QY	1570	GCCGAGAACCGCGAGATCCTCGCGAGCGCGTGCACGCGTGACTACGACCCCGCAGG	1629
DB	461	AlaGluSerArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLys	480
QY	1630	GACCTGTGGCCGAGATCTCAGAAGCAGGGCCACACCGAGTGGACCTACACAGATCTAC	1689
DB	481	AspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnValTyrGln	500
QY	1690	GAGCCCTTCAAGAACCTTGAAGACCGGCAGATGACGCCAAGATGCGCACCGGCCAC	1749
DB	501	GluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsn	520

QY	1750	GACGTGAAGCAGCTGACCGAGGCGCTGCAGAGATCGCCATGAGAGCATCGTGAATCTGG	1809
DB	521	AspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrp	540
QY	1810	GGCAAGACCCCCAAGTTCGCCTGCCCATCCAGAAGGAGACCTGGGAGACCTGGTGGACC	1869
DB	541	GlyLysIleProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpThr	560
QY	1870	GACTACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCGCCCTGGTGTG	1929
DB	561	AspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuVal	580
QY	1930	AAGCTGTGTGTTACAGCTGGAGAGAGCCCATCATCGGCGCCGAGACCTTCTACGTGGAC	1989
DB	581	LysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrTyrTyrValAsp	600
QY	1990	GGCGCCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGCGCG	2049
DB	601	GlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArg	620
QY	2050	CAGAAGATCGTGAGCCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCAG	2109
DB	621	GlnLysIleValThrLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGln	640
QY	2110	CTGGCCCTGCAGGACACGCGCAGCGAGGTGAACATCGTGACCGACCGCATCGCCCTG	2169
DB	641	LeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeu	660
QY	2170	GGCATCATCGGCCCGCAGCCGACAGAGCGAGCGAGCTGTGTGAACCATCATCGAG	2229
DB	661	GlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGlnIleIleGlu	680
QY	2230	CAGCTGATCAAGAGGAGAGAGTGTACTCTGAGCTGGGTGCCGCCCAAGGCGCATCGGC	2289
DB	681	GlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGly	700
QY	2290	GGCAACGAGCAGATCGCAAGCTGGTGAGCAAGGCGCATCCGCAAGGTGTCTTCCTGGAC	2349
DB	701	GlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAsp	720
QY	2350	GGCATCGAT 2358	
DB	721	GlyIleAsp 723	
RESULT 7			
AAE37601			
ID	AAE37601 standard; protein; 998 AA.		
AC	AAE37601;		
XX			
DT	23-OCT-2003 (revised)		
DT	27-AUG-2003 (first entry)		
XX			
DE	HIV-1 subtype C isolate Dul51 reverse transcriptase (RT) protein.		
KW	Regulatory gene; accessory gene; HIV; human immunodeficiency virus;		
KW	vaccine; infection; gene therapy; reverse transcriptase; RT; enzyme.		
OS	Human immunodeficiency virus 1.		
XX			
PN	WO2003037919-A2.		
XX			
PD	08-MAY-2003.		
XX			
PF	31-OCT-2002; 2002WO-IB004550.		
XX			
PR	31-OCT-2001; 2001ZA-00008978.		
XX			
PA	(SAME-) SOUTH AFRICAN MEDICAL RES COUNCIL.		
PA	(UYCA-) UNIV CAPE TOWN.		
XX			
PI	Williamson C, Van Harmelen JH, Gray CM, Bourn W, Karim SA;		
XX			

DR WPI: 2003-430497/40.  
 DR N-PSDB; AAD29258.  
 XX New molecules comprising HIV-1 subtype isolate regulatory/accessory  
 PT genes, useful for manufacturing a vaccine for treating or preventing HIV  
 PT infection.  
 XX Disclosure; Page 94-97; 97pp; English.

XX The invention relates to molecules comprising HIV-1 subtype isolate  
 CC regulatory/accessory genes (tat, nef and rev genes) and modifications and  
 CC derivatives thereof. The invention also provides proteins encoded by such  
 CC genes. Sequences of the invention are useful for manufacturing vaccines  
 CC for treating or preventing human immunodeficiency virus (HIV) infections.  
 CC They are also useful in gene therapy. The present sequence is HIV-1  
 CC subtype C isolate Dnl51 reverse transcriptase (RT) protein. (Updated on  
 CC 23-OCT-2003 to standardise OS field)

XX SQ Sequence 998 AA;

# Alignment Scores:

Pred. No.: 1.4e-195 Length: 998  
 Score: 3527.00 Matches: 662  
 Percent Similarity: 96.51% Conservative: 29  
 Best Local Similarity: 92.46% Mismatches: 21  
 Query Match: 77.43% Indels: 4  
 DB: 6 Gaps: 2

US-09-610-313B-32 (1-2457) x AAE37601 (1-998)

QY 223 TTCCGCGAGGACCTGGCTTCCCGAGGCGAGGCGCGAGTCCCGAGGACGAGAC 282  
 DB 1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20  
 QY 283 CGCCGCAACGCCCCACGAGCGCGAGCTGCAGGTGCGCGCGCAACCCCGCGAGCGAG 342  
 DB 21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgArgAsnAsnProArgSerGlu 40  
 QY 343 GCGCGCGCGAGCGCGAGCGCGACCTGAACCTCCCGAGATCACCTGTGGCAGCGCCC 402  
 DB 41 ThrGlyAlaGluArgLysGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60  
 QY 403 CTGTGTGACATCAGGTGGGGCGGCAGATCAAGAGCGCCCTGTGTGACACCGCGCGAC 462  
 DB 61 LeuValSerIleLysIleGlyGlyGlnThrArgGluAlaLeuLeuAspThrGlyAlaAsp 80  
 QY 463 GACACCGTGTGGAGGAGATGAGCTGCCGCGAAGTGGAGCCCAAGATGATCGCGGC 522  
 DB 81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGly 100  
 QY 523 ATCGCGCGCTTCATCAAGGTGCGCGAGTACGACAGATCTGTGATCGAGATCTGCGCAAG 582  
 DB 101 IleGlyGlyPheIleLysValArgGlnThrAspGlnIleLeuIleGluIleCysGlyLys 120  
 QY 583 AAGCCATCGCACCGTGTGATCGGCCACACCCCGTGAACATCATCGCGCGCAACATG 642  
 DB 121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140  
 QY 643 CTGACCCAGTGGGCTGCACCTCACTCCCATCAGCCCATCGAGCGGTGCCCGTG 702  
 DB 141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160  
 QY 703 AAGCTGAAGCCCGCGAGCGCGCCCAAGGTGAAGCAGTGGCCCTCGACCGAGGAGAG 762  
 DB 161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180  
 QY 763 ATCAAGCCCTGACCGCATCTCGAGGAGATGAGAGGAGGCGCAAGATCAACAAGATC 822  
 DB 181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200  
 QY 823 GGCCCCAGAACCCCTACACACCCCGTGTGCCATCAGAGAGGAGGAGGAGGAGGAG 882  
 DB 201 GlyProGluAsnProThrAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220

QY 883 TGGCGCAAGCTGGTGGAGCTTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTG 942  
 DB 221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240  
 QY 943 CAGCTGGGATCCCGCCACCGCGCGCTGAAGAAGAGAGAGCGTACCGTGTGTGAC 1002  
 DB 241 GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAsp 260  
 QY 1003 GTGGCGCAGCGCTACTTTCAGCGTGGCGCGCTGACAGGAGTTCGCGAAGTACACCCCTTC 1062  
 DB 261 ValGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysTyrThrAlaPhe 280  
 QY 1063 ACCATCCCGAGCATCAACACGAGACCCCGCATCCGCTACCACTCAACAGTGTGTGCC 1122  
 DB 281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuPro 300  
 QY 1123 CAGGCGTGGAGGCGAGCGCGCGCGATCTTCAGAGCAGCATGACCAAGATCTGTGAGGCC 1182  
 DB 301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320  
 QY 1183 TTCCGCGCGCGCAACCCCGAGATCTGTATCTACACAG-----GCCCGCTGTAGCTGGCG 1236  
 DB 321 PheArgAlaGlnAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGly 340  
 QY 1237 AGCGACTGGAGATCGCGCGAGCACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCCTGTG 1296  
 DB 341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluLeuArgGluHisLeuLeu 360  
 QY 1297 CGCTGGGCGCTTCCACCGCGCGACAGAGCACAGAGGAGCGCGCTTCTCTCTG----- 1350  
 DB 361 LysTrpGlyPheThrThrProAspLysHisGlnLysGluProProPheLeuTrpMet 380  
 QY 1351 CCCATCGAGCTGACCCCGCAAGTGGACCTGCGAGCGCCATCGAGCTGCCCGAGAGGAG 1410  
 DB 381 GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400  
 QY 1411 AGCTGGACCGTGAACGACATCCAGAGCTGTGTGGCGAAGCTGAACTGGCGCGAGCCAGATC 1470  
 DB 401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420  
 QY 1471 TACCGCGCGATCAAGGTGCGCGAGCTGTGAGCTGTGCGCGCGCGCGCAAGCGCTGACC 1530  
 DB 421 TyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThr 440  
 QY 1531 GACATCGTGGCCCTGACCGAGGAGCGAGCTGAGCTGCGCGAGAGCCGCGAGATCTGTG 1590  
 DB 441 AspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeu 460  
 QY 1591 CGCGAGCGCGTGCACGCGGTGTACTACGACCCCGAGCAAGGACCTGTGTGGCGAGATCCAG 1650  
 DB 461 LysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGln 480  
 QY 1651 AAGCAGGCGCACGACGAGTGGAGCTTACGAGATCTACGAGAGCGCTTCAAGAACTGAAG 1710  
 DB 481 LysGlnGlyAspAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLys 500  
 QY 1711 ACCGGCAAGTACCGCAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1770  
 DB 501 ThrGlyLysTyrAlaLysArgArgThrHisThrAsnAspValLysGlnLeuThrGlu 520  
 QY 1771 GCCTGTGAGAAGATCGCATCGAGAGCATCGTGTATCTGGCGCGAGAGACCCCAAGTCTCGCG 1830  
 DB 521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540  
 QY 1831 CTGCCCATCCAGAGGAGACCTGGGAGACTGTGGAGCGCTACTACTGCGAGCGCGCGCTGG 1890  
 DB 541 LeuProIleGlnLysGluThrTrpGluIleTrpThrAspTyrTrpGlnAlaThrTrp 560  
 QY 1891 ATCCCGCGAGTGGAGTTCGTGACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1950  
 DB 561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 580

QY 1951 AAGGAGCCCATCATCGCGCGGAGACCTTCTACGTGGACGGCGCGCCAAACGCGAGACC 2010  
 Db LysGluProIleAlaGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThr 600  
 QY 2011 AAGATCGGCAAGCGCGTACGTGACCGACCGGGCGCGGAGAAAGATCGTGAAGCTGACC 2070  
 Db LysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValThrLeuSer 620  
 QY 2071 GAGACCCACCAAGACCGAGCGTGCAGGCGCATCCAGCTGGCCCTGCAGGACAGCGCG 2130  
 Db GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640  
 QY 2131 AGCGAGGTGAACATCGTACCGACGACGAGCGATACGCGCTGGGCGATCATCCAGGCCCGCC 2190  
 Db SerGluValAsnIleIleThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 660  
 QY 2191 GACAAGACGAGGAGCGTGTGTGAACGAGATCATCGAGCGAGCTGATCAAGAAGGAAG 2250  
 Db AspArgSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluArg 680  
 QY 2251 GTGTACCTGAGCTGGTGGTCCCGCCACAAAGGCGATCGCGCGCAACGAGCATCGACAAG 2310  
 Db ValTyrLeuSerTrpValProAlaHisAsnGlyLeuAlaGlyAsnGluHisValAspLys 700  
 QY 2311 CTGGTGAGCAAGGCGATCGCAAGGTGCTTCTCGGACGGCGATCGAT 2358  
 Db LeuValSerArgGlyIleArgLysValLeuValLeuAspGlyIleAsp 716  
 RESULT 8  
 AAB86169  
 ID AAB86169 standard; protein; 3025 AA.  
 XX AC AAB86169;  
 XX DT 11-SEP-2003 (revised)  
 XX DT 21-AUG-2001 (first entry)  
 XX DE HIV-1 subtype C protein fragment #1.  
 XX KW Infection; diagnosis; human; humoral immune response; antiviral;  
 KW cellular immune response; vaccine; treatment; gene therapy.  
 XX OS Human immunodeficiency virus 1.  
 XX FH Key Location/Qualifiers  
 FT Misc-difference 1..3025  
 FT /label= Xaa  
 FT /note= "Xaa represents a stop codon"  
 XX DE10056747-A1.  
 XX PD 31-MAY-2001.  
 XX PF 16-NOV-2000; 2000DB-01056747.  
 XX PR 16-NOV-1999; 99DB-01055089.  
 XX (SHAO/) SHAO Y.  
 PA (GENE-) GENEART GMBH GES ANGEWANDTE BIOTECHNOLOG.  
 XX Wagner R, Wolf H, Shao Y, Graf M;  
 DR WPI: 2001-336417/36.  
 DR N-PSDB; AAH20870.  
 XX New nucleic acid sequences from a human immune deficiency virus  
 PT intersubtype, useful for treatment, prevention and diagnosis of  
 PT infection.  
 XX Disclosure; Fig 8A-O; 48pp; German.  
 PS This invention describes a novel polynucleotide isolated from human  
 CC immunodeficiency virus type 1 subtype C/B' which can be used for the

CC induction of specific humoral and cellular immune responses. (I) and  
 CC polypeptides (II) encoded by them, are useful in pharmaceuticals,  
 CC vaccines and diagnostic agents, particularly for treatment or prevention  
 CC of human immune deficiency virus-1 (HIV-1) infections, also for rational  
 CC design of test or therapeutic reagents, or gene therapy vectors.  
 CC Polypeptides, especially antibodies, specifically directed against (II)  
 CC are similarly useful as pharmaceutical and diagnostic agents. (I) are  
 CC specific for intersubtype C/B' of HIV-1 so are useful in regions  
 CC (particularly China and South-East Asia) where this subtype is prevalent.  
 CC The products of the invention have antiviral activity. This sequence  
 CC represents a protein encoded by the HIV-1 subtype C genome described in  
 CC the method of the invention. (Updated on 11-SEP-2003 to standardise OS  
 CC field)  
 XX SQ Sequence 3025 AA;  
 Alignment Scores:  
 Pred. No.: 4.55e-195 Length: 3025  
 Score: 3520.00 Matches: 678  
 Percent Similarity: 91.14% Conservative: 42  
 Best Local Similarity: 85.82% Mismatches: 59  
 Query Match: 77.28% Indels: 11  
 DB: 4 Gaps: 6  
 US-09-610-313B-32 (1-2457) x AAB86169 (1-3025)  
 QY 15 GGCGGAGCCATGAGCCAGGCGCACGCGCAATCTCTGATGAGCGACGACGACCTTCAA 74  
 Db GlyArgGlyAsnGluProAsnLysGln---CysHisThrAspAlaGluLysGlnPhe\*\*\* 434  
 QY 75 GGGCCCCAAGCGCATCATCAAGTGTCTCAATGGCGCAAGGAGGCGCCACATCGCCCGCAA 134  
 Db ArgLeu\*\*\*LysAsnCys\*\*MetPheGlnLeuTrpGlnGlyArgAlaHisSerGlnLys 454  
 QY 135 CTGCGCGCGCCCGCCGCAAGAGGGCTGTGAAAGTGCAGCAAGGAGGCGCCACGATGAA 194  
 Db LeuGlnGlyPro\*\*\*GluLysGlyLeuLeuGluMetTrpLysArgArgThrProAsnGlu 474  
 QY 195 GGATGTGACCGGCGCGCCAGCACTTCTTCGGGAGGAGCTGGCTTCCCGCAGGGCAA 254  
 Db ArgLeuTyr\*\*\*GluThrGlyGln-PhePheArgGluAsnLeuAlaLeuProGlnGlyAr 494  
 QY 255 GGCGCGCGAGTTCCCGGAGCAGAACCGCGCGCAACAGCCCGCCACCGCGCGAGCTGCA 314  
 Db AlaArgGluPheSerSerGluGlnThrArgAlaAsnSerProThrArgGlyGluLeuGl 514  
 QY 315 GGTG-----CGCGCGCAACACCCCGCGAGCGAGCGCGCGCGCGCGAGGCGCACCT 368  
 Db nValTrpGlyArgAspAsnAsnSerIleSerGluAlaGlyAlaAsnArgGlnGlyThrIl 534  
 QY 369 G-----NACTTCCCGCAGATCACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 422  
 Db eSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThrIleLysIleGl 554  
 QY 423 CGGCGAGATCAAGGAGCGCGTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGAGAT 482  
 Db yGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaGlyAspThrValLeuGluAspLe 574  
 QY 483 GAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCGATCGCGCGCGCTTCATCAAGT 542  
 Db uAsnLeuProGlyLysTrpLysProLysMetIleGlyIleGlyIleGlyIleLysVa 594  
 QY 543 GCGCGAGTACGACGAGTCTGTGATCGATCTGGCGCAAGAGGCGCGCGCGCGCGCGCGT 602  
 Db lArgGlnTyrGluGlnIleProIleGluIleGlyHisLysAlaIleGlyThrVal 614  
 QY 603 GATCG 662  
 Db uValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCysTh 634  
 QY 663 CTGAACCTTCCCGCATCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATGA 722  
 Db rLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAs 654



QY 723 CGGCCCCAAGGTGAAGCAAGTGGCCCTTCACCCGAGGAGAGATCAAGGCCCTTGACCGCCAT 782  
Db 654 pGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaI 674  
QY 783 CTGGGAGAGATGAGAGAGGAGCAAGATCACCAAGATCGCCCGGAGAACCCCTACAA 842  
Db 674 eCyAspGluMetGluLysGluGlyLysIleThrLysIleGlyProGluAenProTyrAS 694  
QY 843 CACCCCTGTTCGTCATCAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 902  
Db 694 nThrProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPh 714  
QY 903 CCGGAGCTGAACAGCGCACCCAGGACTTCTGGAGAGTGACAGTGGGECATCCCCACCC 962  
Db 714 eArgGluLeuAenLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisPr 734  
QY 963 CGCCGCTGAAG 1022  
Db 734 oAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSe 754  
QY 1023 CGTCCCTCGACAGAGACTTCCGCAAGTACACCGCTTCCACCATCCCGACATCAACAA 1082  
Db 754 rIleProLeuThrGluAspPheArgLysTyrThrAlaPheThrIleProSerArgAsnAs 774  
QY 1083 CGAGACCCCGGCTACCGCTACCAAGTACACAGTGTCTGCCCGAGGCTGGAGGCGAGCC 1142  
Db 774 nGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerLe 794  
QY 1143 CAGCATCTTCAGAGCAGCATGACAGATCCTGGAGCCCTTCGCGCCCGCCACCCCGA 1202  
Db 794 uAlaIlePheGlnSerSerMetThrLysThrLeuGluProPheArgLysGlnAsnProGl 814  
QY 1203 GATCGTGATPACAG-----GCCCCCTCTAGCTGGGAGAGAGAGAGAGAGAGAGAG 1256  
Db 814 yIleValIleThrGlnTyrMetAspLeuThrValGlySerAspLeuGluIleGlyL 834  
QY 1257 GCACCGGCCAAGATCGAGGAGCTGCGCAAGCACCCTGCTGCGCTGGGCTTTCACACCCC 1316  
Db 834 nHisArgThrLysIleGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrPr 854  
QY 1317 CGACAAGAGACACAG 1370  
Db 854 oAspLysLysHis---LysGluProProPheLeuThrMetGlyTyrGluLeuHisProAs 873  
QY 1371 CAAGTGAGCGCTGAGCCATCGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1430  
Db 873 pLysTrpThrValGlnProThrGlnLeuProGluLysAspSerTrpThrValAsnAspIl 893  
QY 1431 CCAGAAGCTGTGGGCAAGCTGAACCTGGCCAGCCAGATCTACCCCGGATCAAGGTGCG 1490  
Db 893 eGlnLysLeuValGlyLysLeuAenTrpAlaSerGlnIleTyrProGlyIleLysValAr 913  
QY 1491 CCAGCTGTGAAGCTGTGCGCGGCGCAAGCCCTGACCGACATCTGCGCCCTGACCGCA 1550  
Db 913 gGlnLysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGl 933  
QY 1551 GGAGCGGAGCTGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1610  
Db 933 uGluAlaGluLeuGluLeuAlaGluAenArgGluIleLeuLysGluProValHisGlyVa 953  
QY 1611 GTACTAGACCCAGCAAGAGACTGTGGCGGAGATCCAGAGAGAGAGAGAGAGAGAGAG 1670  
Db 953 lTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnGluGlnTr 973  
QY 1671 GACTTACAGATCTACAGGAGCCCTTCAAGACCTGAAGACCGGCAAGTACGCCAAGAT 1730  
Db 973 pThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMe 993  
QY 1731 CGGACCGCCCGCACACAGAGCTGAAGAGCTGACCGAGGCGCTGACAGAGATCGGCAT 1790  
Db 993 tArgThrAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMe 1013

QY 1791 GGAGAGCATCTGATCTGGGGCAAGAGACCCCAAGTTCCGCTGCCATCCACAGAGAGAC 1850  
Db 1013 tGluGlyIleValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluTh 1033  
QY 1851 CTGGGAGACCTGGGAG 1910  
Db 1033 rTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVa 1053  
QY 1911 GAACACACCCCTGTGTGAAGCTGTGTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1970  
Db 1053 lAsnThrProProLeuValLysLeuTyrTyrGlnLeuGluLysAspProIleValGlyVa 1073  
QY 1971 CGAGACCTTCTACCTGAG 2030  
Db 1073 lGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTy 1093  
QY 2031 CGTCACACAGCGGCGGCGGAGAGAGATCGTGAGAGCTGACCGAGAGAGAGAGAGAGAG 2090  
Db 1093 rValThrAspArgGlyArgLysLysIleValSerLeuThrGluThrThrAsnGlnLysTh 1113  
QY 2091 CGAGCTCAGGCCATTCAGCTGCGCTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2150  
Db 1113 rGluLeuGlnAlaIleCysIleAlaLeuGlnAspSerGlySerGluValAsnIleValTh 1133  
QY 2151 CGACAGCAGTACCGCTTGGGATCATCCAGCCCGAGCCCGAGAGAGAGAGAGAGAGCT 2210  
Db 1133 rAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLe 1153  
QY 2211 GGTGAACAGATCATCAG 2270  
Db 1153 uValAsnGlnIleIleGluGlnLeuMetLysLysGluArgValTyrLeuSerTrpValPr 1173  
QY 2271 CGCCCAAGAGGAGATCGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2330  
Db 1173 oAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleAr 1193  
QY 2331 CAAGGTCTGTCTTCGACCGGATCGAT 2358  
Db 1193 gLysValLeuPheLeuAspGlyIleAsp 1202

RESULT 9  
AAM48949  
ID AAM48949 standard; protein; 854 AA.  
XX  
AC AAM48949;  
XX  
DT 29-AUG-2003 (revised)  
DT 19-APR-2002 (first entry)  
XX  
DE HIV-1 subtype C isolate Dul51 pol protein.  
XX  
KW HIV-1 subtype C; vaccine; HIV infection; AIDS; pol; antiviral.  
XX  
OS Human immunodeficiency virus; type I.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 58  
FT Misc-difference 72 /note= "encoded by TGA"  
FT Misc-difference 72 /note= "encoded by TAA"  
FT Misc-difference 114 /note= "encoded by TAA"  
FT Misc-difference 117 /note= "encoded by TAA"  
XX  
XX WO200204494-A2.  
PD 17-JAN-2002.  
XX  
PF 09-JUL-2001; 2001WO-IB001208.  
XX  
PR 07-JUL-2000; 2000US-0216995P.  
PR 10-JUL-2000; 2000ZA-00003437.

PR 15-SEP-2000; 2000ZA-00004924.  
 XX (MEDI-) MEDICAL RES COUNCIL.  
 PA (UYCA-) UNIV CAPE TOWN.  
 XX (UYNC-) UNIV NORTH CAROLINA.  
 PI Williamson C, Swanstrom RI, Morris L, Karim SA, Johnston RE;  
 XX WPI; 2002-171700/22.  
 DR N-PSDB; AAL41592.  
 XX  
 PT Selecting HIV-1 subtype C isolates, which are useful in developing  
 PT vaccines against HIV infection, comprises isolating viruses with high  
 PT sequence identity to a consensus sequence whose phenotype is associated  
 PT with the HIV subtype.  
 XX  
 PS Claim 19; Page 67; 69pp; English.  
 XX  
 CC The present invention relates to a process for the selection of human  
 CC immunodeficiency virus (HIV) subtype isolates for use in the development  
 CC of a prophylactic and/or therapeutic pharmaceutical composition. The  
 CC process involves selecting isolated virus or viruses with a high sequence  
 CC identity to a consensus sequence and a phenotype which is associated with  
 CC transmission for the particular HIV subtype. The composition can be used  
 CC in the production of vaccines against HIV. The present sequence is the  
 CC HIV-1 subtype C isolate Dui51 pol protein. (Updated on 29-AUG-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 854 AA;  
 Alignment Scores:  
 Pred. No.: 9,24e-193 Length: 854  
 Score: 3478.00 Matches: 662  
 Percent Similarity: 93.77% Conservative: 30  
 Best Local Similarity: 89.70% Mismatches: 31  
 Query Match: 76.36% Indels: 15  
 DB: 5 Gaps: 4  
 US-09-610-313B-32 (1-2457) x AAM48949 (1-854)  
 QY 171 CGGCAAGAGGCGCCACGATGAAGACTGACACCGAGCGCCAGGCCAATCTTCTT----- 224  
 Db 118 ArgGlnGlyPheProSerHisAspValValLysArgArgProValProSerLeuHisAla 137  
 QY 225 -----CCGCGAGGACTGGCC---TTCCCGCCAGGCGAAGCGCCG 260  
 Db 138 CysArgSerThrLeuGluAspProArgValPro-SerSerPheProGlnGlyProAlaAr 157  
 QY 261 CGAGTTCCCGACGAGCAGAACCGCGCCACAGCCCGCCAGCGAGCTGCAGGTGG 320  
 Db 157 GlnPheProSerGlnGlnThrArgAlaAsnSerProThrSerArgGluLeuGlnValAr 177  
 QY 321 CGGCGACAAACCCCGCAGCGCGCGCGCGCCAGCGCCAGGCGCCAGCTTCCCGCA 380  
 Db 177 GargAspAenProArgSerGluThrGlyAlaGluArgLysGlyThrLeuAenPheProGl 197  
 QY 381 GATCACCCTGTGGAGCGCCCTGTGTAGCATCAAGTGTGGCGCGCAGATCAAGAGGC 440  
 Db 197 nileThrLeuTrpGlnArgProLeuValSerileLysileGlyGlyGlnThrArgGluAl 217  
 QY 441 CTTGCTGGACCGCGCGCAGCAGCAGCTGTGTGGAGGAGATGAGCTGCCCGCAAGTG 500  
 Db 217 aleuLeuAenPThrGlyAlaAspAsthValLeuGluAspIleAenLeuProGlyLysTr 237  
 QY 501 GAAGCGCCAGATGATCGCGCGCATCGCGGCTTCATCAAGTGTGGCGCAGTACGACAGAT 560  
 Db 237 pLysProLysMetIleGlyGlyIleGlyGlyPheileLysValArgGlnThrAspGlnI 257  
 QY 561 CTGTGATCGAGATCTCGCGCAAGAGCCATCGGCGCCGCTGTGTGTGATCGCGCCCGCT 620  
 Db 257 eleuileGluileCysGlyLysLysAlaileGlyThrValLeuValGlyProThrProVa 277  
 QY 621 GAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGAACTTCCCCATCAG 680

Db 277 lAsnIleileGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAenPheProIleSe 297  
 QY 681 CCCATCGAGACCGTGCCTGAAAGCTGAAGCCCGGCGATGAGCGCCCAAGGTGAAGCA 740  
 Db 297 rProileGluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGl 317  
 QY 741 GTGGCCCTGACCGAGGAGAGATCAAGCCCTGACCGCCATCTGCGAGGAGATCGAGAA 800  
 Db 317 nTrpProLeuThrGluValLysIleLysAlaLeuThrAlaIleCysGluGluMeGluL 337  
 QY 801 GGAGGCAAGATCAACAGATCGGCCCGGAGAACCCCTTACACACCCCGCTGTCGCCAT 860  
 Db 337 sGluGlyLysIleThrLysIleGlyProGluAenProTyRAsnThrProIlePheAlaI 357  
 QY 861 CAAGAAGAAAGACAGCACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGG 920  
 Db 357 eLysLysGluAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAenLysAr 377  
 QY 921 CACCCAGACTTCTGGGAGGTGCAGCTGGGCGATCCCCACCCCGCGCCCTGAAGAAGAA 980  
 Db 377 gThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLys 397  
 QY 981 GAAGAGCTGACCGTGTGCGACGCGCGCGACGCTTCTTACAGCGTGCCTTGGACAGCA 1040  
 Db 397 sLysSerValThrValLeuAenValGlyAspAlaTyRPheserValProLeuAspGlu 417  
 QY 1041 CTTCCGCAAGTACACCGCTTTCACATCCCGAGCATCAACACGAGACCCCGCGCATCG 1100  
 Db 417 yPheArgLysTyRThrAlaPheThrIleProSerIleAenAenGluThrProGlyIleAr 437  
 QY 1101 CTACAGTACAACTGTCTGCCCGCGGCTGGAAGGCGAGCCCGCATCTTCCAGAGAG 1160  
 Db 437 gTyRGlnTyRAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAlaSe 457  
 QY 1161 CATGACCAAGATCTTGGAGCCCTTCCGCGCGCGCAACCCCGAGATCGTGATCTACAG-- 1218  
 Db 457 rMetThrLysIleLeuGluProPheArgAlaLysAenProGluIleValIleTyRGlnTy 477  
 QY 1219 ----GCCCGCTGTAGTGGGAGCGACTCGAGATCGGCGAGCACCGCGCCCAAGATCGA 1274  
 Db 477 rMetAlaAlaLeuTyRValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGl 497  
 QY 1275 GGAGTGGCGCAAGCACCTGTGCGCTGGGCTTACACACCCCGCAGACAGACAGACAGAA 1334  
 Db 497 uGluLeuArgGluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnL 517  
 QY 1335 GGAGCCCGCTTCTGTG-----CCCATCGAGTGCACCCCGCACAGTGGAGCGCTGAGCC 1388  
 Db 517 sGluProProPheLeuTrpMetGlyTyRLeuLeuHisProAspLysTrpThrValGlnPr 537  
 QY 1389 CATCAGTGTCCCGAAGAGGAGAGCTGACCGGTGAACGACATCCAGAGCTGGTGGGCA 1448  
 Db 537 oileGlnLeuProGluLysAspSerTrpThrValAenAspIleGlnLysLeuValGlyL 557  
 QY 1449 GCTGAGTGGCGCAGCAGATCTACCCCGGCATCAAGTGTGGCGCGCTGTGCAAGCTGT 1508  
 Db 557 sLeuAenTrpThrSerGlnIleTyRProGlyIleLysValArgGlnLeuCysLysLeuLe 577  
 QY 1509 GCGCGCGCCCAAGGCGCTGACCGCATCTGTGCCCTTGCACCGAGGAGGCGCGAGCTGAGCT 1568  
 Db 577 uArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluLe 597  
 QY 1569 GGCCGAGAACCGCGAGATCTCTGCGCGAGCCGCTGTGACCGCGTGTACTACGACCCAGCA 1628  
 Db 597 uAlaGluAenArgGluIleLeuLysGluProValHisGlyValTyRtyrAspProSerL 617  
 QY 1629 GGACCTGTGTGGCGAGATCCAGAGCGGCGCCACGACGAGTGGAGCTACCATCATCCA 1688  
 Db 617 sAspLeuIleAlaGluIleGlnLysGlnGlyAspAspGlnTrpThrTyRGlnIleTyRGl 637  
 QY 1689 GGAGCCCTTCAAGAACCTTGAAGACCGCGCAAGTGTACGCCAAGATGCGCACCGCCACACCA 1748

Db 637 nGluProPheLysAsnLeuLysThrGlyLysTyAlaLysArgThrThrHisThrAs 657  
 Qy 1749 CGACGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGCAGAGCATCGTGATCTG 1808  
 Db 657 nAspValLysGlnLeuThrGluAlaValGlnLysLeuSerLeuGluSerIleValThrTr 677  
 Qy 1809 GGGCAAGACCCCAAGTTCGCTCCGCTCCATCCAGAGAGACCTGGAGACCTGGTGGAC 1868  
 Db 677 pGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluIleTrpTrpTh 697  
 Qy 1869 CGACTACTGGCAGCGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCGCCCTGGT 1928  
 Db 697 rAspTyTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuVal 717  
 Qy 1929 GAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCGCCGAGACCTTCTACGTGA 1988  
 Db 717 lLysLeuTrpTyrglnLeuGluLysGluProIleAlaGlyAlaGluThrPheTyValAs 737  
 Qy 1989 CGGCGCGCCGACCGCAGACCAAGATCGGCAAGCGCGGTACGTGACCGAGCGGCGCG 2048  
 Db 737 pGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyValThrAspArgGlyAr 757  
 Qy 2049 GCAGAAGATCGTGAGCTGACCGAGACCCCAACCAAGAGACCGAGCTGCAGGCCATCCA 2108  
 Db 757 gGlnLysIleValThrLeuSerGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGl 777  
 Qy 2109 GCTGGCCCTCGAGACAGCGCAGCGAGGTGAACATCGTGACCCAGACCAAGTACGCGCT 2168  
 Db 777 nLeuAlaLeuGlnAspSerGluSerGluValAsnIleValThrAspSerGlnTyAlaLe 797  
 Qy 2169 GGGCATCATCAGGCCCGAGCCCGCAAGAGCGAGCGAGCTGGTGAACCAAGATCATCGA 2228  
 Db 797 uGlyIleGlnAlaGlnProAspArgSerGluSerGluLeuValAsnGlnIleIleGl 817  
 Qy 2229 GCAGCTCATCAAGAGAGAGGTGTACCTGAGCTGGTGGCGCGCCCAAGAGCGCATCGG 2288  
 Db 817 uGlnLeuIleLysLysGluArgAlaTyLeuSerTrpValProAlaHisLysGlyIleGl 837  
 Qy 2289 CGGCAACGAGCAGATCAGCAAGCTGTGTGAGCAAGGCGCATCGCAAGGTGTGTG 2340  
 Db 837 yGlyAspGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 854

RESULT 10  
 AAB69282  
 ID AAB69282 standard; protein; 1000 AA.  
 XX  
 AC AAB69282;  
 XX  
 DT 12-SEP-2003 (revised)  
 DT 20-APR-2001 (first entry)  
 XX  
 XX HIV-1 non-subtype B clone 92RW009-6 pol protein.  
 XX  
 DE HIV-1; human immunodeficiency virus; non-subtype B; gag; pol; env; vpu;  
 KW vif; vpr; tat; rev; nef; vaccine.  
 KW  
 XX Human immunodeficiency virus 1.  
 OS  
 XX  
 XX W0200026416-A1.  
 XX  
 XX 11-MAY-2000.  
 PD  
 XX 25-OCT-1999; 99WO-US024837.  
 XX  
 XX 02-NOV-1998; 98US-00184418.  
 PR  
 XX (UABR-) UAB RES FOUND.  
 PA  
 XX Hahn BH, Shaw GM, Gao F;  
 PI  
 XX  
 XX WFI; 2000-365651/31.  
 DR  
 XX Novel genomic nucleic acids of non-subtype B human immunodeficiency virus

PT type 1 useful for detecting and treating AIDS comprises a specific  
 PT nucleotide sequence.

PS Claim 41; Fig 15; 131pp; English.

XX The present in invention provides the protein and coding sequences for a  
 CC number of human immunodeficiency virus (HIV) type 1 non-subtype B  
 CC isolates. The sequences shown include the near full-length coding  
 CC sequences from each isolate, and the env, pol, vif, vpr, vpu, gag, tat,  
 CC rev and nef proteins. These can be used to detect the presence of HIV-1  
 CC in a sample and to produce antibodies against non-subtype B HIV-1 virus.  
 CC These antibodies can be used in vaccines to prevent and treat HIV  
 CC infection. (Updated on 12-SEP-2003 to standardise OS field)

XX Sequence 1000 AA;

Alignment Scores:  
 Pred. No.: 1.52e-192 Length: 1000  
 Score: 3474.50 Matches: 656  
 Percent Similarity: 95.40% Conservative: 29  
 Best Local Similarity: 91.36% Mismatches: 28  
 Query Match: 76.28% Indels: 5  
 DB: 3 Gaps: 3

US-09-610-313B-32 (1-2457) x AAB69282 (1-1000)

Qy 220 TTCTTCGCGCAGACCTGGCTTCCCCAGCGCAAGCCCGCGAGTTCCCGAGCGAGCAG 279

Db 1 PhePheArgGluAsnLeuAlaPheGlnGlnGlyGluAlaArgLysPheSerProGluGln 20

Qy 280 AACCGCGCCCAACGCCACCCAGCGCGAGCTGCAGTGCAGCGGC---GACAAACCCCGC 336

Db 21 ThrGlyAlaAsnSerProThrSerArgGluLeuTrpAsnGlyGlyArgAspSerLeuSer 40

Qy 337 AGCGAGCGCGCGCGCGAGCGCCAGGACACCTGACCTTCCCGAGATCACCTGTGCGAG 396

Db 41 SerGluThrGlyAlaGluArgGlnGlyThrPheAsnPheProGlnIleThrLeuTrpGln 60

Qy 397 CGCCCTCGTGTGAGCATCAAGTGGCGCGCAGATCAAGAGGCGCTGTGACACCGCGC 456

Db 61 ArgProLeuValThrValLysIleGlyGlyGlnLeuArgGluAlaLeuLeuAspThrGly 80

Qy 457 GCCGACGACACCGTGTGTGGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATC 516

Db 81 AlaAspAspThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetile 100

Qy 517 GCGGCGATCGCGCGCTTCATCAAGTGCAGCGCGAGTACACAGATCCTGTGAGATCTGC 576

Db 101 GlyGlyIleGlyGlyPheIleLysValLysGlnTyAspGlnIleLeuIleGluIleCys 120

Qy 577 GCGAAGAGCGCATCGGCACCGTGTGATCGCGCCACCCCGCGTGAACATCATCGCGCGC 636

Db 121 GlyLysLysAlaIleGlyThrValLeuValGlyProThrSerValAsnIleIleGlyArg 140

Qy 637 AACATGCTGACCGAGCTGGCTGCACCTGACCTTCCCATCAGCCCATCGAGACCGTG 696

Db 141 AsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrVal 160

Qy 697 CCGGTGAGCTGAAGCCCGCATGAGCGCGCCAGTGAAGCGAGTGCCTTCCCGAGCGAG 756

Db 161 ProValAlaLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGlu 180

Qy 757 GAGAAGATCAAGCGCCCTGACCGCATCTGCGAGGAGATGAGAGAGAGGGCAAGATCAC 816

Db 181 GluLysIleLysAlaLeuArgGluIleCysThrGluMetGluLysGluGlyLysIleSer 200

Qy 817 AAGATCGCGCCCGAGAACCCCTTACACACCCCGTGTTCGCCCATCAAGAGAGAGCAGC 876

Db 201 LysIleGlyProGluAsnProTyAsnThrProValPheAlaIleLysLysLysAspSer 220

Qy 877 ACCAAGTGGCGCAAGCTGTGTGAGTCTCCGCGAGCTGAACAGCGCACCCAGGACTTCTGG 936

Db 221 ThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrp 240

QY 937 GAGGTGACGTGGGCGATCCCCACCCCGCCGCTGAAGAGAAAGAGCGGTGACCGTG 996  
Db 241 GluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysValThrVal 260  
QY 997 CTGGACGTGGGCGACCGCTACTTCACGCGTCCCGCTGGACGAGACTCCGCAAGTACAC 1056  
Db 261 LeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLysTyrThr 280  
QY 1057 GCCTTCACATCCCGCAGCATCAACACAGACCCCGGCATCCGTCACAGTACACCGTG 1116  
Db 281 AlaPheThrIleProSerIleAsnAsnGlnThrProGlyIleArgTyrGlnTyrAsnVal 300  
QY 1117 CTGCCCCAGGCGTGAAGGGACGCCAGCATCTTCAGAGCAGCATCAACAGATCCCTG 1176  
Db 301 LeuProGlnGlyTrpLysGlySerProAlaIlePheGlnAsnSerMetThrLysIleLeu 320  
QY 1177 GAGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCTACCCAG- - - - -GCCCCCTGTAC 1230  
Db 321 GluProPheArgAlaGlnAsnGlnGluIleValIleTyrGlnTyrMetAspAspLeuTyr 340  
QY 1231 GTGGCAGCAGCTCGAGATCGGCAGCACCGCCGCAAGATCGAGAGCTGCGCAAGCAC 1290  
Db 341 ValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGlnGluLeuArgGluHis 360  
QY 1291 CTGCTCGCTGGGCTTCACACCCCGCACAGAGCAGCAGAGAGCCCGCTTCCTG 1350  
Db 361 LeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeu 380  
QY 1351 - - - - -CCCATCGAGCTGCACCCCGACCAAGTGGACCGTGCAGCCCATCGAGCTGCCGAG 1404  
Db 381 TrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGlu 400  
QY 1405 AAGGAGAGCTGACCGTGACGATCCAGAACGCTGGTGGGCAAGCTGAACCTGGCCAGC 1464  
Db 401 LysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSer 420  
QY 1465 CAGATCTACCCCGCATCAGGTGGCGCAGCTGTCAGAGCTGCTGGCGGCGCCAGGCC 1524  
Db 421 GlnIleTyrProGlyValLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAla 440  
QY 1525 CTGACCGACATCGTCCCTCGACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCGAG 1584  
Db 441 LeuThrAspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGlu 460  
QY 1585 ATCTGCGCGAGCCCGTGCACGGCTGTACTACGACCCCGCAGCAAGACTGTGTGCCGAG 1644  
Db 461 IleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGlu 480  
QY 1645 ATCCAGAGCAGGCGCCAGCCAGTGGACCTACAGATCTACCGAGGCCCTTCAGAAC 1704  
Db 481 IleGlnLysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsn 500  
QY 1705 CTGAAGACCGCAAGTACGCCAAGATGGCGACCGCCACCAACGAGCTGAAGCAGCTG 1764  
Db 501 LeuLysThrGlyLysTyrAlaLysArgThrAlaHisThrAsnAspValLysGlnLeu 520  
QY 1765 ACCGAGGCGGTGAGAGATCCCATGGAGAGCATCGTGATCTGGGCAAGACCCCAAG 1824  
Db 521 ThrGluAlaValGlnLysIleAlaMetGluSerIleValIleTrpGlyLysThrProLys 540  
QY 1825 TTCGCGCTGCCCATCCAGAGAGACCTGGGAGACTGTGGACCGACTACTGGCAGGCC 1884  
Db 541 PheArgLeuProIleGlnLysGluThrTrpGluThrTrpThrAspTyrTrpGlnAla 560  
QY 1885 ACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCTCGTGAAGCTGTGGTACCAAG 1944  
Db 561 ThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGln 580  
QY 1945 CTGGAGAGGAGCCCATCATCCTGGCGCCGAGACTTCTACGTGGACGGCGCCCAACCGC 2004  
Db 581 LeuGluLysGluProIleLeuGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArg 600

QY 2005 GAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGCGGCGCGGAGATCGTGAGC 2064  
Db 601 GluThrLysIleGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSer 620  
QY 2065 CTGACCGAGACCCCAACCAAGACCGAGGTGAGGCCATCCAGCTGGCCCTGCAGGAC 2124  
Db 621 LeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAsp 640  
QY 2125 AGCGCAGCAGGTGAACATCGTACCGACAGCAGCAGTACGCCCTGGGCGATCCAGGCC 2184  
Db 641 SerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAla 660  
QY 2185 CAGCCGCAAGAGCAGAGCGAGGTGTGAACCAAGATCATCGACGAGCTGATCAAGAAG 2244  
Db 661 GlnProAspSerSerGluSerGluAlaValAsnGlnIleIleGluGlnLeuIleLysLys 680  
QY 2245 GAGAAGGTGTACCTGAGCTGGTGGTCCGCCCAAGGGCATCGCGCGCAACGAGCAGATC 2304  
Db 681 GluArgValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnVal 700  
QY 2305 GACAAGCTGTGAGCAGAGCGCATCGGCAAGGTGCTGTCTCGACGGCGCATCGAT 2358  
Db 701 AspLysLeuValSerSerGlyIleArgArgValLeuPheLeuAspGlyIleAsp 718  
RESULT 11  
AAW72993  
ID AAW72993 standard; protein; 1002 AA.  
XX  
AC AAW72993;  
XX  
DT 25-MAR-2003 (revised)  
DT 15-FEB-1999 (first entry)  
XX  
XX HIV isolate LAV.MAL pol protein.  
XX  
XX LAV.MAL; HIV; human immunodeficiency virus; diagnosis; vaccine; AIDS;  
KW pol.  
XX  
XX Human T-lymphotropic virus.  
XX  
XX US9824482-A.  
XX  
XX 20-OCT-1998.  
XX  
XX 06-JUN-1995; 95US-00471474.  
XX  
XX 23-JUN-1986; 86FR-00040138.  
PR 13-APR-1987; 87US-00038330.  
PR 19-FEB-1991; 91US-00656797.  
PR 10-DEC-1992; 92US-00988530.  
PR 18-NOV-1993; 93US-00154397.  
XX  
XX (INSP ) INST PASTEUR.  
XX  
XX Wain-Hobson S, Sonigo P, Alizon M, Montagnier L;  
XX  
XX WPI; 1998-582548/49.  
DR N-P5DB; AAV63467.  
XX  
XX Human immunodeficiency virus isolate LAV (MAL) - and method for detecting  
PT anti-HIV antibodies.  
XX  
XX Disclosure; Fig 7B-E; 47pp; English.  
XX  
CC This is the amino acid sequence of the pol protein of lymphadenopathy  
CC associated virus LAV.MAL (NCIM 1-641), a new virus isolate from Zaïre  
CC that is responsible for diseases clinically related to AIDS. The sequence  
CC was deduced from an open reading frame (ORF) of the LAV.MAL genome (see  
CC AAV63467). 7 ORFs (see AAW72992-98) were identified. Specific peptides of  
CC the envelope glycoprotein can be used as antigens in a claimed method for  
CC the in vitro detection of an antibody directed against LAV. The method is  
CC useful for the diagnosis of AIDS or pre-AIDS, or to detect antibodies in  
CC patients, asymptomatic carriers and in blood-related products. LAV viral

CC antigens are also useful in vaccines. (Updated on 25-MAR-2003 to correct  
CC PR field.)

XX  
SQ Sequence 1002 AA;

Alignment Scores:  
Pred. No.: 1,68e-191 Length: 1002  
Score: 3456.50 Matches: 649  
Percent Similarity: 95.56% Conservative: 39  
Best Local Similarity: 90.14% Mismatches: 25  
Query Match: 75.88% Indels: 7  
DB: 2 Gaps: 4

US-09-610-313B-32 (1-2457) x AA#W2993 (1-1002)

QY 220 TTCTTCGGGAGGACCTGGCTTCCCGGAGGAGGCGCGGAGTTCCTCCCGAGGAGCAG 279  
DB 1 PhePheArgGluAenLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20  
QY 280 AACCGCGCAACAGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 336  
DB 21 ThrArgAlaAenSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40  
QY 337 AGCGAGGCGCGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAG 390  
DB 41 SerGluThrGlyAlaGluArgGlnGlyValSerPheSerPheProGlnThrLeu 60  
QY 391 TGGCAGGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 450  
DB 61 TrpGlnArgProValThrValArgValGlyGlnLeuLysGluAlaLeuLeuAsp 80  
QY 451 ACCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 510  
DB 81 ThrGlyAlaAspAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys 100  
QY 511 ATGATCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 570  
DB 101 MetileGlyGlyGlyGlyGlyPheileLysValArgGlnTrpAspGlnThrLeuLeuGlu 120  
QY 571 ATCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 630  
DB 121 IleCysGlyLysLysAlaileGlyThrileuValGlyProThrProValAsnileile 140  
QY 631 GCGCGCAACATGTCAGCCAGCTGGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCTG 690  
DB 141 GlyArgAsnMetLeuThrGlnileGlyCysThrLeuAsnPheProileSerProileGlu 160  
QY 691 ACCGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 750  
DB 161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180  
QY 751 ACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 810  
DB 181 ThrGluGluLysileLysAlaLeuThrGluileCysLysAspMetGluLysGluGlyLys 200  
QY 811 ATCACCAGATCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 870  
DB 201 IleLeuLysileGlyProGluAenProTyAsnThrProValPheAlaileLysLysLys 220  
QY 871 GACAGCACCAAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 930  
DB 221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240  
QY 931 TTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 990  
DB 241 PheTrpGluValGlnLeuGlyileProHisProAlaGlyLeuLysLysLysSerVal 260  
QY 991 ACCGTGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1050  
DB 261 ThrValLeuAspValGlyAspAlaTyThrPheSerValProLeuAspGluAspPheArgLys 280  
QY 1051 TACACGCGCTTACCATTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1110

DB 281 TyrThrAlaPheThrileProSerileAsnAsnGluThrProGlyLysArgTyrGlnTyr 300  
QY 1111 AACGTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1170  
DB 301 AsnValLeuProGlnGlyTrpLysGlySerProAlailePheGlnSerSerMetThrLys 320  
QY 1171 ATCTGAGGCGCTTCCCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1224  
DB 321 IleLeuGluProPheArgThrLysAsnProGluileValileTyrGlnTyrMetAspAsp 340  
QY 1225 CTGTACGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1284  
DB 341 LeuTyrrValGlySerAspLeuLysileGlnHisArgThrLysileGluLysLysLys 360  
QY 1285 AAGCACCTGCTGCGGCGGCTTCCACACCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1344  
DB 361 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
QY 1345 TTCTGTG-----CCCATCGAGCTGCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1398  
DB 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProileGlnLeu 400  
QY 1399 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1458  
DB 401 ProAspLysGluSerTrpThrValAsnAspLysGlnLysLysLysLysLysLysLys 420  
QY 1459 GCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1518  
DB 421 AlaserGlnileTyrProGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 440  
QY 1519 AAGGCGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1578  
DB 441 LysAlaLeuThrAspLysValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 460  
QY 1579 CGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1638  
DB 461 ArgGlnileLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 480  
QY 1639 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1698  
DB 481 AlaGluileGlnLysGlnGlyGlnTrpThrTrpThrGlnileTyrGlnGlnTyr 500  
QY 1699 AAGAACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1758  
DB 501 LysAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 520  
QY 1759 CAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1818  
DB 521 GlnLeuThrGluAlaValGlnLysileAlaGlnGluSerileValileTrpGlyLysThr 540  
QY 1819 CCCAAGTTCCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1878  
DB 541 ProLysPheArgLeuProLysLysLysLysLysLysLysLysLysLysLysLysLysLys 560  
QY 1879 CAGGCGCACTGGATCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1938  
DB 561 GlnAlaThrTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 580  
QY 1939 TACGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1998  
DB 581 TyrGlnLeuGluThrGluProileValileGlyAlaGluThrPheTyrrValAspGlyAlaAla 600  
QY 1999 AACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2058  
DB 601 AsnArgGluThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 620  
QY 2059 GTGAGGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2118  
DB 621 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaileHisLeuAlaLeu 640  
QY 2119 CAGGAG 2178  
DB 641 GlnAspSerGlySerGluValAsnileValThrAspSerGlnTyrAlaLeuGlyLysile 660

QY 2179 CAGGCCAGCCGACAGAGCGAGCGAGCTGGTGAACACGATCATCGAGCAGTGTATC 2238  
Db 661 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGluGlnLeuIle 680  
QY 2239 AAGAGGAGAGAGTGTACTGAGCTGGGTGCCGCCACAGAGGCGATCGGGGCAACGAG 2298  
Db 681 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 700  
QY 2299 CAGATCGACAAGCTGGTGAAGGCGATCCGCAAGGTGCTGTTCTCTGGACGGCATCCAT 2358  
Db 701 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

## RESULT 12

ID AAO30963 standard; protein; 1003 AA.

AC AAO30963;

DT 06-OCT-2003 (first entry)

DE HIV pol protein.

XX Human immunodeficiency virus; HIV; pol; IV9; immunotherapy; infection;  
KW epitope; vaccine; enzyme.

XX Human immunodeficiency virus.

XX WO2003051285-A2.

XX 26-JUN-2003.

PF 29-OCT-2002; 2002WO-US034620.

PR 29-OCT-2001; 2001US-0345116P.

PA (GENZ ) GENZYME CORP.

PA (GENO ) GEN HOSPITAL CORP DBA MASSACHUSETTS GEN.

XX Nicolette CA, Walker BD;

XX WPI; 2003-541603/51.

PT Novel synthetic therapeutic peptide compounds useful for inducing an  
PT immune response in a mammal, specifically cross-react with HIV IV9  
PT epitopes.

PS Claim 8; Page 65-67; 67pp; English.

CC The invention relates to synthetic therapeutic peptide compounds that  
CC specifically cross-react with HIV (human immunodeficiency virus) pol  
CC protein IV9 epitopes. The invention is useful for treating HIV infection.  
CC The invention is also used in immunotherapy and as vaccines. The present  
CC sequence is HIV pol protein

SQ Sequence 1003 AA;

## Alignment Scores:

Pred. No.:	2,346-191	Length:	1003
Score:	3454.00	Matches:	651
Percent Similarity:	95.01%	Conservative:	34
Best Local Similarity:	90.23%	Mismatches:	28
Query Match:	75.83%	Indels:	8
DB:	6	Gaps:	4

US-09-610-313B-32 (1-2457) x AAO30963 (1-1003)

QY 220 TTCCTCCGAGGAGCTGGCTTCCCCAGGCGAGGCCGAGTTCCTCCAGCGAGGAG 279  
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
QY 280 AACCCGCGCAACAGCCCAACAGCGCGAGCTGCAGGTG-----CGCGGCGACAACCCC 333  
|||||

Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTTrpGlyArgAspAsnAsnSer 40  
QY 334 CGCAGCGAGCGCGCGCGCGAGCGCGAGCGCGACCTG-----AACTTCCCGCAGATCACC 387  
Db 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60  
QY 388 CTGTGGCAGCGCCCGCTGGTGAGCATCAAGTGGCGGCCAGATCAAGAGGCGCCCTGTGTG 447  
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
QY 448 GACACCGCGCGCGAGCAGCACCGCTGCTGAGGAGATGAGCTGCCCGCAAGTGGAGAGCCC 507  
Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100  
QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGGCGCCAGTACGACCATCTGTATC 567  
Db 101 LysMetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIle 120  
QY 568 GAGATCTCGCGCAAGAAGGCCATCGGCACCTGCTGATCGCGCCCGCCCGCTGAACATC 627  
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
QY 628 ATCGCGCCCAACATCTGACCCAGCTGGGTGTCACCTGAATCTCCCATCAGCCCATC 687  
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
QY 688 GAGACCGTCCCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCGCCAGGTGAAGCAGTGGCCC 747  
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACGCCATCTCGGAGGAGATGGAGAGGAGGCG 807  
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
QY 808 AAGATCAACAAGATCGCGCCCGGAGAACCCCTACAAACCCCGCTGTTCCCATCAAGAAG 867  
Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGAATCTCCCGAGCTGAAACAAGGCGACCCAG 927  
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
QY 928 GACTTCTCGGAGGTGACGTGGCATCCCGCCCGCGCGCTCAAGAAGAAGAAGAGAGC 987  
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
QY 988 GTGACCGTGTGCGAGCGCGACGCTTCTACGCTGCCCTGAGCGAGGAGTCTCCGC 1047  
Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg 280  
QY 1048 AAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTACCCAG 1107  
Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
QY 1108 TACAACGTGTGCCCGCGCGCTGGAAAGGCGAGCCCGACGCTTCTCCAGAGCAGCATGACC 1167  
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
QY 1168 AAGATCTCTGGAGCCCTTCGCGCGCGCAACCCCGAGATCGTATCTACAG-----GCC 1221  
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340  
QY 1222 CCCCTGTACGTGGCGAGCAGCTGGAGATCGGCGAGCAGCCCGCCCAAGATCGAGGAGCTG 1281  
Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
QY 1282 CGCAGACACCTGTCTCGCTGGCGCTTCACACCCCGCGCAAGAAGCACCAAGAGGAGCC 1341  
Db 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380  
QY 1342 CCCTTCCTCTG-----CCCATCGAGCTGCACCCCGCAAGAGTGAGACCGTGAGCCCATCGAG 1395  
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
|||||

```
Qy 1396 CTGCCCGAGAGAGAGTGGACCGTGAACGACATCCAGAGCTGGTGGCAAGCTGAAC 1455
Dy 1401 LeuProGluLysAspSerTrpThrValAenAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy 1456 TGGGCGCAGCTACCCCGGATCAAGTGGCGGCGGCTGCAAGCTCTGGCGGC 1515
Dy 421 TrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuCysLysLeuLeuArgGly 440
Qy 1516 GCCAAGGCCCTGACCGACATCGTCCCTGACCGAGGCGGAGCTGGAGCTGGCCGAG 1575
Dy 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
Qy 1576 AACCGGAGATCTCGCGGAGCCCGTGCACCGGCTGTACTACGACCCCGAGCAAGACTG 1635
Dy 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
Qy 1636 GTGCGCGAGATCCAGAGCGGCGCACGACAGTGGACCTACAGACTACAGAGCCG 1695
Dy 481 IleAlaGluIleGlnLysGlnGlyGlnTyrThrTyrGlnIleTyrGlnGluPro 500
Qy 1696 TTCAGAACCTGAAGACCGGAGTACCGCAAGATGCGCACCGCCACCAACGAGCTG 1755
Dy 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520
Qy 1756 AAGCAGCTGACCGAGCGCTGCAGAAATCCCATGGAGAGCATCGTGATCTGGGCAAG 1815
Dy 521 LysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIleValIleTyrGlyLys 540
Qy 1816 ACCCGCAGTTCGCGCTGCCATCCAGAGAGAGCTGGGAGACTGGTGACCGACTAC 1875
Dy 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
Qy 1876 TGGCAGGCGACTCGATCCCGAGTGGGAGTTCGTGAACACCCCGCTGGTCAAGCTG 1935
Dy 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
Qy 1936 TGGTACCAGCTGGAGAGGAGCCCATCATCGCGCCGAGACTTCTACGTGGAGCGGCGC 1995
Dy 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
Qy 1996 GCCAACCGCAGACCAAGATCGGAGCGCGGTACGTGACCGAGCGGCGCGCGCAAG 2055
Dy 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
Qy 2056 ATCTGTAGCTGACCGAGACCAACCAACAGAGACCGAGCTGCAGGCCATCCACTGCGC 2115
Dy 621 ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640
Qy 2116 CTGCAGGACACGCGCAGAGGTGAACATCGTGACCGACCGAGCTGACGCGCTGGGCATC 2175
Dy 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy 2176 ATCCAGCGCCAGCCGACAGAGCGAGAGCGAGCTGGTGAACACGAGCATCGACAGCTG 2235
Dy 661 IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleLeuGluLeu 680
Qy 2236 ATCAAGAGGAGAGGTGTACTGAGCTGGTGGCGCCCGCCACAGAGGCGATCGGCGCAAC 2295
Dy 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700
Qy 2296 GAGCAGATCCAGACAGCTGGTGAGCAAGGCGATCCGCAAGGTGTGTCTGGAGCGGCATC 2355
Dy 701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
Qy 2356 GAT 2358
Dy 721 Asp 721
```

RESULT 13

ADN36413

ID ADN36413 standard; protein; 1003 AA.

XX

```
AC ADN36413;
XX 15-JUL-2004 (first entry)
XX HIV protein for anti-HIV vaccine.
XX anti-HIV; vaccine; HIV; promoter; viral particle; immunization.
XX Human immunodeficiency virus 1.
XX WO2004035006-A2.
XX 29-APR-2004.
XX 17-OCT-2003; 2003WO-US033112.
XX 18-OCT-2002; 2002US-0419465P.
XX (AARO-) AARON DIAMOND AIDS RES CENT.
XX Huang Y, Ho DD, Chen Z;
XX WPI; 2004-348328/32.
XX N-PSDB; ADN36412.
XX Nucleic acid vector comprising at least one HIV sequence operably linked
XX to a promoter and encoding a protein that does not assemble into viral
XX particles, useful in immunizing a subject against HIV infection.
XX Disclosure; SEQ ID NO 27; 166pp; English.
XX The invention relates to a nucleic acid vector comprising at least one
XX HIV sequence operably linked to a promoter and encoding a protein that
XX does not assemble into viral particles. The nucleic acid vector is useful
XX in immunizing a subject against HIV infection. This sequence corresponds
XX to a nucleic acid used in the invention.
XX SQ Sequence 1003 AA;
XX Alignment Scores:
XX Pred. No.: 2,34e-191 Length: 1003
XX Score: 3454.00 Matches: 651
XX Percent Similarity: 95.01% Conservative: 34
XX Best Local Similarity: 90.29% Mismatches: 28
XX Query Match: 75.83% Indels: 8
XX DB: 8 Gaps: 4
XX US-09-610-313B-32 (1-2457) x ADN36413 (1-1003)
```

```
Qy 220 TTCTTCGCGAGGACCTGGGCTTCCCGCAGGCAAGGCCGCGAGTTCCCGCAGCGAGCAG 279
Dy 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
Qy 280 AACCGCGCCAAACGCCCCACCGAGCGCGAGCTGCAGGTG-----CGCGGCGCAACACCC 333
Dy 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnSer 40
Qy 334 CGCAGCGAGCGCGCGCGCGAGCGCGAGCGCACCTG-----ACTTCCCGCAGTACC 387
Dy 41 ProSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnValThr 60
Qy 388 CTGTGGCAGCGCGCGCGCGCGAGCATCAAGTGGCGCGCGCAGATCAAGAGGCGCGCTGTG 447
Dy 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCGCGAGCACCGCTGCTGGAGAGATGAGCTGCGCGCAAGTGAAGCC 507
Dy 81 AspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100
Qy 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGTGGCGCGAGTACGACCGAGTCTGATC 567
Dy 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTrpAspGlnIleLeuLeu 120
```



Qy	568	GAGATCTCGCGCAAGAAGGCATCGGCACCGTGTCTGATCGCGCCACCCCGCGTGAACATC	627
Db	121	GlulileCysGlyHisIleLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle	140
Qy	628	ATCGCGCGCAACATGCTGACCCAGCTCGGCTGCACCTGAACCTTCCCATCAGCCCATC	687
Db	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
Qy	688	GAGACCGTGCCTGGAAGCTCAAGCCCGGCATGACGCGCCCAAGGTGAAGCAGTGGCCC	747
Db	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTTPPro	180
Qy	748	CTGACCGAGGAGATCAAGCGCTGACCGCCATCTGCAGGAGATGGAAGAAGGAGGCG	807
Db	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
Qy	808	AAGATCACCAAGATCGCGCCCGAGAACCCCTACAAACCCCGCTGTTCGCCATCAAGAG	867
Db	201	LysIleSerLysIleGlyProGluAsnProGlyAsnThrProValPheAlaIleLysLys	220
Qy	868	AAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGAGCTTGAACAGCCACCCAG	927
Db	221	LysAspSerThrLysTTPArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln	240
Qy	928	GACTTCTGGGAGTGACGTGGGCATCCCCACCCCGCCCTGAAGAAGAAGAAGAGC	987
Db	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
Qy	988	GTGACCGTGTGACGTGGGAGCGCCTACTTACGCGTGCCTCTGGACGAGGACTTCCGC	1047
Db	261	ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArg	280
Qy	1048	AAGTACACCGCTTACCATCCCAGCATCAACAAACGAGACCCCGCGCATCCGCTACCAG	1107
Db	281	LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln	300
Qy	1108	TACAACTGTCTGCCCGCAGGGCTGGAAGGAGCGCCACAGCATCTTCCAGAGCATCATGAC	1167
Db	301	TyrAsnValLeuProGlnGlyTyrLysGlySerProAlaIlePheGlnSerSerMetThr	320
Qy	1168	AAGATCTGGAGCCCTTCGGGCGCGCAACCCCGAGATCTGTATCTACAG-----GCC	1221
Db	321	LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp	340
Qy	1222	CCCTGTGATGTGGCAGCGACTTGGAGATCGGCCAGCACCGCCCAAGATCGCAGGAGCTG	1281
Db	341	AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu	360
Qy	1282	CGCAAGCACCTGTCTGCGCTGGGCTTCCACACCCCGCCCAAGAAGCACCAAGAAGGAGCC	1341
Db	361	ArgGlnHisLeuLeuAspGlyLeuThrThrProAspLysLysHisGlnLysGluPro	380
Qy	1342	CCCTTCTCTG-----CCCATCGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAG	1395
Db	381	ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal	400
Qy	1396	CTGCGCCGAGGAGAGCTGCACGCTGAACGACATCCAGAGCTGTGTGGCGCAGCTGAAC	1455
Db	401	LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
Qy	1456	TGGCGCCAGCAGATCTACCCCGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCG	1515
Db	421	TrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly	440
Qy	1516	GCCAAAGCCCTGACCGACATCGTGCCCTGACCGAGGAGCGCAGCTGTGAGCTGGCGCAG	1575
Db	441	ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu	460
Qy	1576	AACCGCGAGATCTTCGCGGAGCCGCTGCACGGCGGTGTACTACGACCCACGACGAGCTG	1635
Db	461	AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu	480
Qy	1636	GTGCGCCGAGATCCAGAAGCAGGGCCACGACAGTGGACCTTACACAGATCTTACACGAGCC	1695

Db	481	ileAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro	500
QY	1696	TTCAAGAACCTTGAGACCCGCAAGTACGCCAAAGATGCGCACCGCCACCAACACGACGTG	1755
Db	501	PhelysasnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlahisThrAsnAspVal	520
QY	1756	AAGCAGCTGACCGAGCCGCTGCAGAAAGATCCGCAATGGAGAGCATCGTGATCTCGGGCAAG	1815
Db	521	LysGlnLeuThrGluAlaValGlnLysIleThrGluSerIleValIleTyrGlyLys	540
QY	1816	ACCCCAAGTTCGCGCTGCCCATCCAGAAAGAGACCTGGGAGACCTGGTGAGCGGACTAC	1875
Db	541	ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr	560
QY	1876	TGGCAGAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTGGTGAAGCTG	1935
Db	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProLeuValLysLeu	580
QY	1936	TGTTACCAAGTGGAGAGGAGCCCATCATCGGGCCGAGACCTTCTACGTGGACGCGCC	1995
Db	581	TrpTyrGlnLeuLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla	600
QY	1996	GCCAAACCGGAGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGCCCGGAGAAG	2055
Db	601	AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys	620
QY	2056	ATCGTGAGCTGACCGAGACCAACCAAGACCGAGCTGCAGGCGCATCCAGCTGGCC	2115
Db	621	ValValThrLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla	640
QY	2116	CTGCAGACAGCGGACGAGGTGAAATCATCGTGACCGACAGCCAGTACGCCCTGGGCATC	2175
Db	641	LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle	660
QY	2176	ATTCAGGCCACCCCGCAAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCAGCTG	2235
Db	661	IleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsnGlnIleIleGluGlnLeu	680
QY	2236	ATCAAGAGGAGGAGGTGATCTACCTGAGCTGGTGCGCGCCACCAAGGCGCATCGCGGCAAC	2295
Db	681	IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLys	700
QY	2296	GAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGCAAGGTGCTGTTCTCTGGACGGCATC	2355
Db	701	GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle	720
QY	2356	GAT 2358	
Db	721	Asp 721	
RESULT 14			
AAP81861			
ID	AAP81861 standard; protein; 1002 AA.		
XX			
AC	AAP81861;		
XX			
DT	27-AUG-2003 (revised)		
DT	25-MAR-2003 (revised)		
DT	16-DEC-1990 (first entry)		
XX			
DE	Sequence encoded by LAV MA L POL gene.		
XX	HIV; HTLV III; AIDS; diagnosis; vaccine; probe; hybridisation.		
XX			
OS	Lymphadenopathy-associated virus.		
XX			
PN	W08707906-A.		
XX			
PD	30-DEC-1987.		
XX			
PF	22-JUN-1987; 87WO-EP000326.		
XX			

```

PR 23-JUN-1986; 86EP-00401380.
XX (INSP ) INST PASTEUR.
XX
XX Allison M, Sonigo P, Wainhobson S, Montagnier L, Alison M;
XX Sonigo P, Wainhobson S, Montagnier L;
XX
XX WPI; 1988-014396/02.
XX N-PSDB; AAN80437.
XX
XX New variants of lymphadenopathy associated virus (LAV) - used for prodn.
XX of DNA, antigens and antibodies used in diagnosis of AIDS and pre-AIDS.
XX
XX Claim 8; Fig 8A-8I; 72pp; English.
XX
XX LAV EL I (AAN80436) and LAV MA L (AAN80437) were isolated from the
XX peripheral blood lymphocytes of patients. Different AIDS virus isolates
XX concerned are designated by 3 letters of the patients name. Stable probes
XX including the DNA sequences can be used for detection of the new LAV
XX viruses or related viruses or DNA proviruses in eg. biological samples.
XX The proteins or peptides can be used for detection of antibodies induced
XX in vivo and present in biological fluids. The DNA can also be used for
XX the expression of LAV viral antigens for the prodn. of a vaccine against
XX LAV. The polypeptides can also be used for the prodn. of antibodies for
XX the detection of proteins related to the LAV viruses, partic. for
XX diagnosis of AIDS or pre-AIDS. (Updated on 25-MAR-2003 to correct PI
XX field.) (Updated on 27-AUG-2003 to correct OS field.)
XX
XX Sequence 1002 AA;

Alignment Scores:
Pred. No.: 3,26e-191 Length: 1002
Score: 3451.50 Matches: 648
Percent Similarity: 95.42% Conservative: 39
Best Local Similarity: 90.00% Mismatches: 26
Query Match: 75.77% Indels: 7
DB: 1 Gaps: 4

US-09-610-313B-32 (1-2457) x AAP81861 (1-1002)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGGAGCCGCGAGTTCCTCCCGAGCGAG 279
DB 1 PhePheArgGluAsnLeuAlaPheProGlnGlyLysAlaArgGluPheProSerGluGln 20
QY 280 AACCGCGCCACACGCCCCAGCCAGCGCGAGCTGCGAGTG---CGCGCGCAACACCCCGC 336
DB 21 ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeu 40
QY 337 AGCAGGCGCGCGCGCGAGCGCCAGGCGC-----ACCTGGAATTCCTCCCGAGATCACCTG 390
DB 41 SerGluThrGlyAlaGluArgGlnGlyLeValSerPheSerPheProGlnLeuThrLeu 60
QY 391 TGGCAGCGCCCTTGGTGAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCGCTCTGAGC 450
DB 61 TrpGlnArgProValThrValArgValGlyGlyGlnLeuLysGluAlaLeuAsp 80
QY 451 ACCGCGCGCGAGACACCGTCTGAGAGAGATGAGCTGCCCGGCAAGTGAAGCCCAAG 510
DB 81 ThrGlyAlaAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys 100
QY 511 ATGATCGCGCGCATCGCGCGCTTCAATCAAGTGGCGCGAGTACGACAGATCCTGATCGAG 570
DB 101 MetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuLeuGlu 120
QY 571 ATCTGCGCGAGAGGCGCATCGGACCGTGTGATCGGCGCGCGCGCGCGCGCGCGCATCATC 630
DB 121 IleCysGlyLysLysAlaIleGlyThrIleLeuValGlyProThrProValAsnIle 140
QY 631 GCGCGCAACATGTGACCGAGCTGGCTGCACTTCCCTGCACTTCCCGCATCAGCCCATCGAG 690
DB 141 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160
QY 691 ACCGTGCGCGTGAAGCTGAAGCCCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCTG 750

161 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 180
751 ACCGAGGAGAGATCAAGGCCCTGACCGCCATCTCGAGAGGAGATCGAGAGGAGGCGCAAG 810
181 ThrGluGlyLysIleLysAlaLeuThrGluIleCysLysAspMetGluLysGluGlyLys 200
811 ATCACAAGATCGCGCGCGCGAGAACCCCTTACAAACACCCCGCTGTTTCGCCATCAAGAAGAG 870
201 IleLeuLysIleGlyProGluAsnProTrpAsnThrProValPheAlaIleLysLysLys 220
871 GACAGACCAAGTGGCGCGAGCTGTCGAGCTTCGCGAGCTGAAACAGCGCACCCAGGAC 930
221 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 240
931 TTCTGGGAGGTGCGAGCTGGCGCATCCCCACCCCGCGCGCTTGAAGAGAAAGAGAGCGCTG 990
241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysLys 260
991 ACCGTGCTGGACGTGGCGCGAGCCCTTACTTACGCGTGGCCCTCGAGCGAGGACTTCGCGCAAG 1050
261 ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArgLys 280
1051 TAGACCCCTTCAACATCCCGAGCATCAACAGAGACCCCGCGGATCCGCTACCGATAC 1110
281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyLysArgTyrGlnTyr 300
1111 AACGTGCTGCGCGCGAGGTGGAAGCGCGCGCGAGCATCTTCCAGAGCAGCATGACCAAG 1170
301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerMetThrLys 320
1171 ATCTGGAGGCGCTTCCCGCGCGCGCAACCCCGAGATCGTGATCTACCAAG-----GCCGCC 1224
321 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 340
1225 CTGTACGTGGCGCGAGACCTGGAGATCGCGCAGCACCGCGCGCAAGATCGAGAGCTGCGC 1284
341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360
1285 AAGCACCTGTGCGTGGCGCTTCCACCCCGCGCGCGCGAGAGACCAAGAGAGAGAGAGAGAG 1344
361 GlnHisLeuLysTrpGlyPheThrProAspLysLysHisGlnLysGluProPro 380
1345 TTCTCTG-----CCATCGAGCTGCGCGCGCGCGAGTGGAGCGCGCGCGCGCGCGCTG 1398
381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 400
1399 CCGAGAGAGAGAGAGCTGGAACCGTGAACGATCAGAACGCTGGTGGGCAAGCTGAACCTGG 1458
401 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420
1459 GCGAGCGAGATCTACCCCGCGCATCAGTGGCGCGCGCTGCGAGCTGCGAGCTGCTGCGCGCG 1518
421 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 440
1519 AAGCGCTGACCGACATCGTGGCGCTTCCACCGCGCGCGCGAGAGCGCGAGCTGGCGCGAG 1578
441 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuProGluAsn 460
1579 CCGAGATCTCTGCGCGAGCGCGCTGCGAGCGGTGTACTACGACCCCGAGAGAGCTGGT 1638
461 ArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeuIle 480
1639 GCGAGATCTCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 1698
481 AlaGluIleGlnLysGlnGlyGlnGlnTrpThrTyrGlnIleTyrGlnGluGlnTyr 500
1699 AAGAACTGAAGCGCGCGAGTACGCCCAAGATGGCGCGCGCGCGCGCGCGCGCGCGCGAG 1758
501 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 520
1759 CAGCTGACCGAGCGCGTGCAGAGAGATCGCATGCGAGAGCATCGTGATCTGGGCGCAAGC 1818

```



201 LysIleSerIysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220 Db  
868 AAGCAGACCAAGTGGCGCAAGCTGGTGAAGTTCGCGAGCTTGAACAAGCGCACCCAG 927 Qy  
221 LysAspSerThrLysTyrArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240 Db  
928 GACTTCTGGGAGGTGCAGCTGGGATCCCCACCCCGCGGCTGAGAGAAAGAGC 987 Qy  
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260 Db  
988 GTGACCGTCTGGAGCTGGGCGAGCGCTTACTTTCAGCTGCCCTGGAGGAGCTCCGC 1047 Qy  
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280 Db  
1048 AAGTACACCGCTTACCACATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTACCAG 1107 Qy  
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300 Db  
1108 TAGAACGTGTCGCCCGAGGCTGAAGGGCAGCCCGAGCATCTTCAGAGCAGCATGACC 1167 Qy  
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320 Db  
1168 AAGATCTCTGGAGCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCCAG-----GCC 1221 Qy  
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyrGlnTyrMetAsp 340 Db  
1222 CCCCTGTACGTGGCGAGCGACTCGAGATCGCCGAGCAGCGCGCAAGATCGAGAGCTG 1281 Qy  
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360 Db  
1282 CGCAAGCACCTGCTGCGCTGGGGCTTCAACACCCCGAGCAAGAGCCACGAGAGGCC 1341 Qy  
361 ArgGlnHisLeuLeuLysTrpGlyPheThrProAspLysLysHisGlnLysGluPro 380 Db  
1342 CCCTTCCTG-----CCCATCGAGCTGCACCCCGCAAGTGGAGCGGCGAGCCCATCGAG 1395 Qy  
381 ProIleLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400 Db  
1396 CTGCCCGAGAGGAGAGCTGGACCGTGAACACATCCAGAGCTGGTGGGCAAGCTGAAC 1455 Qy  
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420 Db  
1456 TGGCCAGCCAGATCTACCCCGGATCAAGTGGCGCAGCTGTGCAAGCTGCTGCGCGGC 1515 Qy  
421 TrpAlaSerGlnIleTyrAlaGlyLysValArgGlnLeuCysLysLeuLeuArgGly 440 Db  
1516 GCCAAGGCCCTGACGACATGTCCTGACCGAGGCGCGAGCTGGAGCTGGCGCGAG 1575 Qy  
441 ThrLysAlaLeuThrGluValIleProLeuThrGluAlaGluLeuLeuAlaGlu 460 Db  
1576 AACCGCAGATCTCTGCGCAGCGCTGCAGCGCGTGTACTACGACCCCGAGAGGACCTG 1635 Qy  
461 AsnArgGluLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480 Db  
1636 GTGCGCAGATCCAGAGAGGCGCACGACAGTGGACCTTACCAGATCTTACCAGAGCCCC 1695 Qy  
481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500 Db  
1696 TTCAGAACCTGAGACGCGCAGTACGCCAGATGGCGACCCCGCCACACCAAGCAGCTG 1755 Qy  
501 PheLysAsnLeuThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspVal 520 Db  
1756 AAGCAGCTGACCGAGGCGGTGCAGAGATCGCCATCGAGAGCATCGTGATCTGGGGCAAG 1815 Qy  
521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540 Db  
1816 ACCCCCAAGTTCGCCCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTAC 1875 Qy  
541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluAlaTyrTrpThrGluTyr 560 Db  
1876 TGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCGCCCTGGTGGAGCTG 1935 Qy  
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580 Db

1936 TGGTACCAGCTGGAGAAAGGAGCCCATCATCGGCCCGAGACCTTCTACGTGAGCGGCC 1995 Qy  
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600 Db  
1996 GCCAACCGCAGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGCGCGCAGAG 2055 Qy  
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLys 620 Db  
2056 ATCGTGAAGCTGACCGAGACCAACACAGAACCGAGCTGCAGGCCATCCAGCTGGCC 2115 Qy  
621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640 Db  
2116 CTCAGGACAGCGCGCAGCGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175 Qy  
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660 Db  
2176 ATCCAGGCCAGCGCCGACAGAGCGAGCGAGCTGTGAACACAGATCATCGACAGCTG 2235 Qy  
661 IleGlnAlaGlnProAspLysSerGluLeuValSerGlnIleIleGluGlnLeu 680 Db  
2236 ATCAAGAGGAGAGGTGTACCTGAGCTGGTGGCCGCCACAGGGCATCGCGCGCAAC 2295 Qy  
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyGlyAsn 700 Db  
2296 GAGCAGATCGACAGCTGGTGAAGCGCATCCGCAAGGTGTCTTCTTGACCGCATC 2355 Qy  
701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720 Db  
2356 GAT 2358 Qy  
721 Asp 721 Db

Search completed: June 2, 2005, 04:43:53  
Job time : 233.424 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 02:53:39 ; Search time 40.7339 Seconds  
(without alignments)

11607.276 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 4555

Sequence: 1 gtcagccaccatg9ccga.....gggctagcaccggtgaattc 2457

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgm2\_1/USPTO\_spool/US09610313/runat\_31052005\_155137\_15159/app\_query.fasta\_1.7893  
-DB=PIR\_79 -QFWT=fastan -SURFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09610313 @CGN 1 1 189 @runat\_31052005\_155137\_15159 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_79:\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3462.5	76.0	1002	2	554378 pol polyprotein -
2	3454.5	75.8	1002	1	GNLJND HIV-1 retropepsin
3	3449	75.3	1003	1	GNVWIV HIV-1 retropepsin
4	3432	75.3	1015	1	GNVWJ3 HIV-1 retropepsin
5	3410	74.9	1003	2	T09440 pol polyprotein -
6	3407	74.8	1003	1	B44001 HIV-1 retropepsin
7	3406	74.8	1003	1	GNVWAZ HIV-1 retropepsin
8	3396	74.6	1012	1	GNVWVL HIV-1 retropepsin
9	3168	69.5	1027	1	GNLJST HIV-1 retropepsin
10	3064	67.3	912	2	S33980 pol polyprotein -
11	3051	67.0	902	2	T01668 pol polyprotein -
12	2728	59.9	559	2	B47175 reverse transcript
13	2708	59.5	559	2	A47175 reverse transcript
14	2307	50.6	1039	2	S46347 pol polyprotein -

15	2170	47.6	1055	2	553092 pol polyprotein -
16	2166.5	47.6	1055	1	GNLJST HIV-1 retropepsin
17	2158	47.4	1036	1	GNLJG2 HIV-1 retropepsin
18	2157.5	47.4	1056	1	GNLJG3 HIV-1 retropepsin
19	2149.5	47.2	1035	1	GNLJGG HIV-1 retropepsin
20	2149	47.2	1034	1	GNLJCA HIV-1 retropepsin
21	2143.5	47.1	1019	2	T11560 pol polyprotein -
22	2130.5	46.8	1009	2	S28081 pol polyprotein -
23	2111.5	46.4	1032	2	S12153 pol polyprotein -
24	2034	44.7	1058	2	S08436 pol polyprotein -
25	2015	44.2	1061	1	GNLJG4 HIV-1 retropepsin
26	1980.5	43.5	1054	1	GNLJG5 HIV-1 retropepsin
27	1925.5	42.3	656	2	S30484 pol polyprotein -
28	1896.5	41.6	656	2	S30483 pol polyprotein -
29	1350	29.6	1124	2	GNLJFP HIV-1 retropepsin
30	1347	29.6	1124	1	GNLJPP HIV-1 retropepsin
31	1305	28.6	1124	2	B45557 HIV-1 retropepsin
32	1122.5	24.6	1109	1	B45345 pol protein - Maed
33	1106.5	24.3	1087	2	JQ1162 HIV-1 retropepsin
34	1102.5	24.2	1086	1	B46335 HIV-1 retropepsin
35	1100	24.1	1146	1	GNLJEW HIV-1 retropepsin
36	1100	24.1	1146	1	GNLJEW HIV-1 retropepsin
37	1095	24.0	1101	1	B45390 HIV-1 retropepsin
38	1091	24.0	1101	1	GNLJVS HIV-1 retropepsin
39	1080.5	23.7	1145	1	GNLJEV HIV-1 retropepsin
40	1071	23.5	219	2	S32152 RNA-directed DNA p
41	1067	23.4	219	2	S32118 RNA-directed DNA p
42	1067	23.4	219	2	S32139 RNA-directed DNA p
43	1065	23.4	219	2	S32089 RNA-directed DNA p
44	1065	23.4	219	2	S32160 RNA-directed DNA p
45	1064	23.4	219	2	S32133 RNA-directed DNA p

#### ALIGNMENTS

##### RESULT 1

S54378

pol polyprotein - human immunodeficiency virus type 1 (fragment)

C:Species: human immunodeficiency virus type 1, HIV-1

C>Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004

C/Accession: S54378

R:Theodore, T.; Buckler-White, A.J.

submitted to the EMBL Data Library, July 1989

A:Reference number: S54377

A:Accession: S54378

A>Status: preliminary

A:Molecule type: genomic RNA

A:Residues: 1-1002 <THE>

A/Cross-references: UNIPROT:P12499; EMBL:M22639; NID:G329377; PIDN:AAA5366.1; PID:G3293

C:Superfamily: pol polyprotein

C:Keywords: polyprotein

Alignment Scores:			
Pred. No.:	3.6e-154	Length:	1002
Score:	3462.50	Matches:	651
Percent Similarity:	95.14%	Conservative:	34
Best Local Similarity:	90.42%	Mismatches:	28
Query Match:	76.02%	Indels:	7
DB:	2	Gaps:	4

US-09-610-313B-32 (1-2457) x S54378 (1-1002)

Qy	220	TTCTTCGCGAGGACCTGGCTTCCCGGAGGCGCGAGTTCCTCCCGAGGAGCAG 279
Db	1	PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaGlyGluLeuSerSerGluGln 20
Qy	280	AACCGGCCAACAGCCCGAGCGAGCTGCGAGTGGCGGC---GACAAACCCCGC 336
Db	21	ThrArgAlaAsnSerProThrSerArgGluLeuArgValTrpGlyArgAspAsnProLeu 40
Qy	337	AGCGAGCGCGCGCGAGCGCGAGCGACCTGACCTTC-----CCCCAGATCACCTG 390
Db	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheAsnCysProGlnIleThrLeu 60

QY 391 TGGCAGCGCCCTGCTGAGCATCAAGTGGCGCCAGATCAAGAGGAGCCCTGCTGGAC 450  
 Db 61 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 80  
 QY 451 ACCGGCGCCGACACACCGTGTGGAGGAGATGAGCCTGCCGGCAAGTGAAGCCCAAG 510  
 Db 81 ThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysProLys 100  
 QY 511 ATGATCGGGGATCGGGCTTCATCAAGGTGGCCAGTAGCAGACAGATCCTGATCGAG 570  
 Db 101 MetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 120  
 QY 571 ATCTCGGCACAAGSGCCATCGCACCGTGTGATCGCGCCACCCCGTGAACATCATC 630  
 Db 121 IleCysGlyHisLeuAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 140  
 QY 631 GCCCGCAACATCTGACCCAGCTGGGCTGCACCTGAACTCCCATCAGCCCAATCGAG 690  
 Db 141 GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 160  
 QY 691 ACCGTGCGCGTGAAGTGAAGCCCGCATGAGCGGCCCAAGTGAAGCAGTGGCCCTG 750  
 Db 161 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu 180  
 QY 751 ACCGAGGAGAGATCAAGCGCCCTGACCGCATCTCGAGGAGATCGAGAGGAGGCAAG 810  
 Db 181 ThrGluGluLysIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys 200  
 QY 811 ATCACAAGATCGGCGCCGAGAACCCCTACAAACCCCGCTTCGCCATCAAGAAGAAG 870  
 Db 201 IleSerArgValGlyProGluAsnProTyrAsnThrProIlePheAlaIleLysLysLys 220  
 QY 871 GACAGACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAACAAGCGCAACCCAGGAC 930  
 Db 221 AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 240  
 QY 931 TTCTGGAGGTGACGTGGCATCCCCACCCCGCGCTGCTGAAGAAGAAGACGCTG 990  
 Db 241 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 260  
 QY 991 ACCGTGCTGAGCGTGGCGACCCCTACTTTCAGCGTGGCCCTCGAGAGGACTTCGCGAAG 1050  
 Db 261 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArgLys 280  
 QY 1051 TACACCGCTTCACCATCCCCAGCATCAACAACGAGACCCCGGATCCGCTACCGATAC 1110  
 Db 281 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 300  
 QY 1111 AACGTGCTGCCCCAGGCTGGAGGGCAGCCCGACATCTTCAGAGCAGCATGACCAAG 1170  
 Db 301 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 320  
 QY 1171 ATCTCGGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATCTACAG-----GCCGCC 1224  
 Db 321 IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrGlnTyrMetAspAsp 340  
 QY 1225 CTGTACGTGGGAGCGACTGAGATTCGGCCAGCACCGCGCCCAAGATCGAGAGTGGCG 1284  
 Db 341 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 360  
 QY 1285 AAGCACCTGCTGGCTGGGCTTCACCAACCCCGCAAGACCAAGCAGGAGGAGCCCCC 1344  
 Db 361 GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 380  
 QY 1345 TTCTCTG-----CCCATCGAGTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGTGTG 1398  
 Db 381 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnSerIleLysLeu 400  
 QY 1399 CCGGAGAGGAGGAGTGGACCGTGAACGATCCAGAGAGCTGTGGGCAAGCTGAACTGG 1458  
 Db 401 ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 420

QY 1459 GCCACGACAGATCTACCCCGCATCAAGGTGGCGCCAGCTGTGCAAGCTGTGCGCGCGCC 1518  
 Db 421 AlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGlyThr 440  
 QY 1519 AAGGCCCTGACGACATCGTCCCTGACCCAGAGAGCGCGAGCTGGAGCTGGCCGAGAAC 1578  
 Db 441 LysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsn 460  
 QY 1579 CGCGAGATCCTCGCGAGCCGCTGACCGCTGTACTAGCAGCCCGCAGCAAGACCTGGTG 1638  
 Db 461 ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 480  
 QY 1639 GCCAGATTCACAGACGAGCGCACACAGCTGGAGCTTACCAGATCTACCAGAGCCCTTC 1698  
 Db 481 AlaGluIleGlnLysGlnGlyHisGlyGlnTrpThrTyrGlnIleTyrGlnGluProPhe 500  
 QY 1699 AAGAACCTGAAGCCGCAAGTACGCCAAGATGGCACCGCCCGCCACCAACAGCACTGAAG 1758  
 Db 501 LysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAlaHisThrAsnAspValLys 520  
 QY 1759 CAGCTGACCGAGCGCTGCAGAGATCGCCATGGAGAGCATCGTGTGGGCGAAGACC 1818  
 Db 521 GlnLeuAlaGluValValGlnLysIleSerThrGluSerIleValIleTrpGlyLysThr 540  
 QY 1819 CCCAAGTTCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGACCGACTACTGG 1878  
 Db 541 ProLysPheArgLeuProIleGlnLysGluThrTrpGluThrTrpTrpValGluTyrTrp 560  
 QY 1879 CAGCGACCTCGATCCCGAGTGGAGTTCGTGAACACCCCCCTCGTGAAGCTGTGG 1938  
 Db 561 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 580  
 QY 1939 TACCACTGGAGAGAGCCCATCATCGCGCGCGAGACCTTCTAGCTGAGCGCGCCGCC 1998  
 Db 581 TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAlaAla 600  
 QY 1999 AACCCGAGACCAAGATCGCAAGCCGCTACGTGACGACCGCGCGCGCGCAGAGATC 2058  
 Db 601 AsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 620  
 QY 2059 GTGAGCTGACCGAGACCAACCAAGACCGAGTTCGAGCGCATCCAGCTGGCCCTG 2118  
 Db 621 ValProPheThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleAsnLeuAlaLeu 640  
 QY 2119 CAGGACAGCGCAGCGAGTGAACATCGTACCGACAGCGAGTACCGCTGGCGATCATC 2178  
 Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 660  
 QY 2179 CAGGCGCCAGCCGACAGAGCGAGCGAGCTGGTGAACACAGATCATCGAGCGCTGATC 2238  
 Db 661 GlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeuIle 680  
 QY 2239 AAGAGAGAGAAGTGTACTGAGCTGGTGGTCCCGCCCAAGGGCATCGCGCGCAACGAG 2298  
 Db 681 LysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsnGlu 700  
 QY 2299 CAGATCGACAGCTGTGAGCAAGGGCATCGCAGAGTGTCTTCTCGACCGCATCGAT 2358  
 Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 2

GNLJND

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate NDK)  
 N:Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
 C:Species: human immunodeficiency virus type 1, HIV-1  
 A:Note: host Homo sapiens (man)  
 C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
 C:Accession: JQ0067  
 R:Spire, B.; Sire, J.; Zachar, V.; Rey, F.; Barre-Sinoussi, F.; Galibert, F.; Hampe, A.,  
 Gene 81, 275-284, 1989  
 A:Title: Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immuno  
 A:Reference number: JQ0065; MUID:90034200; PMID:2806917  
 A:Accession: JQ0067



A:Molecule type: DNA  
 A:Residues: 1-1002 <SPI>  
 A:Cross-references: UNIPROT:P18802; GB:M27323; NID:G328154; PID:AAA44869.1; PID:G328158  
 C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
 C:Genetics:  
 A:Gene: pol  
 C:Superfamily: pol polypeptide  
 C:Keywords: ARDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
 F:56-154/Product: retropepsin #status predicted <RTP>  
 F:80/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
 Pred. No.: 8.49e-154 Length: 1002  
 Score: 3454.50 Matches: 651  
 Percent Similarity: 95.14% Conservative: 34  
 Best Local Similarity: 90.42% Mismatches: 28  
 Query Match: 75.84% Indels: 7  
 DB: 1 Gaps: 4  
 US-09-610-313B-32 (1-2457) x GNLJND (1-1002)

Qy	220	TTCTTCGCGAGGACCTGCGCTTCCCGAGGCGAGCCGCGAGTTCCCGAGCGAGCAG	279
Db			
	1	PhePheArgGluAspLeuAlaPheProGlnGlyysAlaGlyGluPheSerSerGluGln	20
Qy	280	AACCGCGCAACAGCCCGAGCGCGAGCTGCAGGTG--CGCGGCGCAACCCCGCGC	336
Db			
	21	ThrArgAlaAenSerProThrSerArgGluLeuArgValTrpGlyGlyAspAsnProLeu	40
Qy	337	AGCGAGCGCGCGAGCGCGAGCGCGAGCCCTG-----ACTTCCCGCAGATCACCTG	390
Db			
	41	SerGluThrGlyAlaGluArgGlnGlyThrValSerPheSerPheProGlnLeuThrLeu	60
Qy	391	TGGCAGCGCCCTCGTGAGCATCAAGGTGGCGCGCAGATCAAGGAGCGCCCTCTGCAC	450
Db			
	61	TrpGlnArgProLeuValThrLeuLysIleGlyGlnLeuLysGluAlaLeuLeuAsp	80
Qy	451	ACCGCGCGCAGCACCGCTGCGAGGAGATGAGCTCGCCCGCGCAAGTGAAGCCCAAG	510
Db			
	81	ThrGlyAlaAspAspThrValLeuGluGluLeuAsnLeuProGlyLysTrpLysProLys	100
Qy	511	ATGATCGCGCGCATCGCGGCTTATCATGAGTGGCGCGAGTACGACGAGTCTGATCGAG	570
Db			
	101	MetileGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnLeuLeuIleGlu	120
Qy	571	ATCTGCGCAAGAGCGCATCGGCACCTGCTGATCGSCCCACCCCGTGAACATCATC	630
Db			
	121	IleCysGlyTyrlsAlaMetGlyThrValLeuValGlyProThrProValAsnIlelle	140
Qy	631	GGCGCGCAACATGTCAGCCAGCTGGCTGCAACCTGAACTTCCCGCATCGCCCATCGAG	690
Db			
	141	GlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu	160
Qy	691	ACCGTGGCGGTGAGTGAAGCCCGCATGACCGCGCCCAAGGTGAAGTGAAGTGGCCCTG	750
Db			
	161	ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeu	180
Qy	751	ACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAGGAGGGCAAG	810
Db			
	181	ThrGluGlyIleLysAlaLeuThrGluIleCysThrGluMetGluLysGluGlyLys	200
Qy	811	ATCACCAAGATCGGCGCGAGAACCCCTTACAAACACCCCGCTGTTCGCGCATCAAGAGAAG	870
Db			
	201	IleSerArgIleGlyProGluAsnProTyraAsnThrProIlePheAlaIleLysLysLys	220
Qy	871	GACAGCAACAGTGGCGCAGCTGGTGAATTCGCGAGCTGAACAGCGCACCCAGGAC	930
Db			
	221	AspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp	240
Qy	931	TTCTGGGAGGTGAGCTGGGATCCCCACCCCGCGCTGAAGAGAGAGAGAGCGTG	990
Db			
	241	PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal	260

Qy	991	ACCGTGTGAGCTGGGCGAGCCCTACTTACGCTGCGCCCTGGAGGAGACTTCGCGAAG	1050
Db			
	261	ThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArgLys	280
Qy	1051	TACACCGCTTACCATCCCGCAGCATCAACACGAGACCCCGCGCATCGCTACCACTAC	1110
Db			
	281	TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrlnTrp	300
Qy	1111	AACGTGTGCGCCCGAGGCTGGAAGCGGAGCCCGCAGCATCTTCCAGAGCAGCATGACCAAG	1170
Db			
	301	AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys	320
Qy	1171	ATCTGAGAGCTTCCGCGCGCCCAACCCCGAGATGCTATACCAAG-----GCCGCC	1224
Db			
	321	IleLeuGluProPheArgLysGlnAsnProGluIleValIleTyrlnTrpMetAspAsp	340
Qy	1225	CTGTACCTGGGCGAGCTGGAGATCGGCGAGCACCGCGCAAGATCGAGAGCTGGCG	1284
Db			
	341	LeuTyrlValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg	360
Qy	1285	AAGCACCTGTGCTGGGCTTCCACCAACCCCGAGAGACCAAGAGAGAGAGAGAGAG	1344
Db			
	361	GluHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro	380
Qy	1345	TTCTCTG-----CCCATCGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAGTG	1398
Db			
	381	PheLeuTrpMetGlyTyrlGluLeuHisProAspLysTrpThrValGlnProIleAsnLeu	400
Qy	1399	CCCGAGAGGAGAGCTGGACCGTGAAACGACATCCAGAGCTGGTGGGCAAGCTGACTGG	1458
Db			
	401	ProGluLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp	420
Qy	1459	GCCAGCCAGATCTACCCCGCGCATCAAGTGGCGCAGCTGTGCAAGCTGCTCGCGGCGCC	1518
Db			
	421	AlaSerGlnIleTyrlAlaGlyIleLysValLysGlnLeuLysLysLeuLeuArgGlyThr	440
Qy	1519	AAGCCCTGACCGACATCGTGCCTCGACCGAGAGGCGCGAGCTGGAGCTGGCCGAGAAC	1578
Db			
	441	LysAlaLeuThrGluValValProLeuThrGluGluAlaGluLeuLeuAlaGluAsn	460
Qy	1579	CGCGAGATCTGCGCGAGCCCGTGCACGGGTGTACTACGACCCCGAGCAAGGAGCTGTG	1638
Db			
	461	ArgGluIleLeuLysGluProValHisGlyValTyrlAspProSerLysAspLeuIle	480
Qy	1639	GCCGAGATCCAGAGAGGCGCACGACAGTGGACCTTACAGATCTTACAGAGGCGCTTC	1698
Db			
	481	AlaGluLeuGlnLysGlnGlyAspGlyGlnTrpThrTyrlGlnIleTyrlGlnLysProPhe	500
Qy	1699	AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGAGCTGAAG	1758
Db			
	501	LysAsnLeuLysThrGlyLysTyrlAlaArgThrArgGlyAlaHisThrAsnAspValLys	520
Qy	1759	CAGCTGACGAGCGCGTGCAGAGATCGCATCGAGAGCATCGTGTGCTGGGGCAAGACC	1818
Db			
	521	GlnLeuThrGluAlaValGlnLysIleAlaThrLysSerIleValIleTrpGlyLysThr	540
Qy	1819	CCCAAGTTCCTGCGCTCCATCCAGAGAGACCTGGCGAGACCTGGTGACCACTACTGG	1878
Db			
	541	ProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpIleGluTyrlTrp	560
Qy	1879	CAGGCCACTGGATCCCGGAGTGGAGTTCTGTGAACACCCCGCTGGTGAAGCTGTGG	1938
Db			
	561	GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp	580
Qy	1939	TACGAGTGAAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTGGAGCGGCGCGCC	1998
Db			
	581	TyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrlValAspGlyAlaAla	600
Qy	1999	AACCGGAGAGACCAAGATCGGCGCGGCTTACGTGACCGAGCCGCGCGCGCAGCAAGATC	2058
Db			
	601	AsnArgGluThrLysLeuGlyLysAlaGlyTyrlValThrAspArgGlyLysGlnLysVal	620
Qy	2059	GTGAGCCTGACCGAGAGACCAACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGGCCCTG	2118

```
|||||
621 ValProPheThrAspThrThrAsnGlnYsThrGluLeuGlnAlaIleAsnLeuAlaLeu 640
QY 2119 CAGGACAGCGGAGGAGTGTAACATCTGTACCGACAGCAGCAGTACGCGCTGGGCAATC 2178
Db 641 GlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyAlaLeuGlyIleIle 660
QY 2179 CAGGCCCCAGCCGACAGAGCGAGAGCTGGTGAACACAGATCATCGAGCAGCTGATC 2238
Db 661 GlnAlaGlnProAspYserGlnSerGluLeuValSerGlnIleIleGlnGlnLeuIle 680
QY 2239 AAGAAGGAGAAAGTGTACCTGAGCTGGTGCCTCCGCGCCACAGGGCATCGGCGCAACGAG 2298
Db 681 LysLysGluYsValTyLeuAlaIleValProAlaHisLysGlyIleGlyGlyAsnGlu 700
QY 2299 CAGATCGACAGCTGGTGAAGGCGCATCCCAAGGTGCTGTTCTTGGACGGCATCGAT 2358
Db 701 GlnValAspLysLeuValSerGlnGlyIleArgLysValLeuPheLeuAspGlyIleAsp 720

RESULT 3
GNVWLTV
HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate LAV-1a)
N:Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly
C:Species: human immunodeficiency virus type 1, HIV-1
A:Note: host Homo sapiens (man)
C:Date: 17-May-1985 #sequence_revision 17-May-1985 #text_change 03-Jun-2002
C:Accession: A03966
R:Wain-Hobson, S.; Sonigo, P.; Danos, O.; Cole, S.; Alizon, M.
Cell 40, 9-17, 1985
A:Title: Nucleotide sequence of the AIDS virus, LAV.
A:Reference number: A90866; MUID:85099333; PMID:2981635
A:Accession: A03966
A:Molecule type: DNA
A:Residues: 1-1003 <WAI>
C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein
C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle
F:57-155/Product: retropepsin #status predicted <RT>
F:81/Active site: Asp (shared with dimeric partner) #status experimental

Alignment Scores:
Pred. No.: 1 53e-153 Length: 1003
Score: 3449.00 Matches: 651
Percent Similarity: 94.8% Conservative: 33
Best Local Similarity: 90.2% Mismatches: 29
Query Match: 75.72% Indels: 8
DB: 1 Gaps: 4

US-09-610-313B-32 (1-2457) x GNVWLTV (1-1003)
QY 220 TTCTTCGCGAGACCTGGCTCCCTCCCAAGGCAAGCGCCGAGTCCCGCAGGACGAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCCAAAGCCCAAGCCAGCCGCGAGCTGCGAGTG-----CGCGGCGACAACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyArgAspAsnAsnSer 40
QY 334 CGCAGCGAGCGCGCGCGAGCGCCAGCGCCACCCCTG-----AACTTCCCGCAGATCACC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCTGGTGTAGCATCAAGTGGCGCGCCAGATCAAGGAGCCCTGTCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGACACCGCTGTGGAGGAGATGAGCTCCCGCGCAAGTGGAGGCC 507
Db 81 AspThrGlyAlaAspThrValLeuGluGluMetSerLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGGCGCATCGCGCGCTTCATCAAGGTGGCGCAGTACGACCCAGATCCTGATC 567
```

```
|||||
101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyAspGlnIleLeuIle 120
QY 568 GAGATCTCGCGCAGAGAGCGCATCGGCACCGTGTGTGATCGGCCCGCCACCCCGTGAACATC 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
QY 628 ATCGCGCCCAACATGCTGACCCAGCTGGGCTGCACCTGAACCTCCCATCAGAGCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
QY 688 GAGACCGTGCCTGGTGAAGTGAAGCCCGGCATGAGCGCCCAAGGTGAAGCAGTGGGCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCCCATCTCTCGAGGAGATGAGAGAGGAGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAAGATCGCGCCCGAGAACCCCTACACACCCCGCTGTCGCCATCAAGAG 867
Db 201 LysIleSerLysIleGlyProGluAsnProTyArgAsnThrProValPheAlaIleLysLys 220
QY 868 AAGACACGACCAAGTGGCGCAAGCTGTGTGACTTCCGCGAGCTGAAACAAGCGCACCCAG 927
Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240
QY 928 GACTTCTGGAGGTGACGCTGGGCGATCCCCACCCCGCGCGCTGAAAGAAGAGAGAGC 987
Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260
QY 988 GTGACCGTGTGGAGCTGGCGAGCGCTTCTACGCTGCCCTCGAGCGAGGACTTCCGC 1047
Db 261 ValThrValLeuAspValGlyAspAlaTyPheSerValProLeuAspGluAspPheArg 280
QY 1048 AAGTACACCGCTTCCATCCATCCCGCATCAACACAGACACCCCGCGCATCCGCTAC 1107
Db 281 LysTyThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGln 300
QY 1108 TACAACTGTGTCGCCAGGCTGGAGGGAGCGCCAGCATCTTCCAGAGCGATGATACC 1167
Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTCGAGACCTTCCGCGCCGCAACCCCGAGATCGTGATCTACAG-----GCC 1221
Db 321 LysIleLeuGluProPheArgLysGlnAsnProAspIleValIleTyGlnTyMetAsp 340
QY 1222 CCCTGTGACGTGGCGAGCTGAGATCGGCGAGCATCCGCGCAGCACCCGCGCAAGATCGAGGAGCTG 1281
Db 341 AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360
QY 1282 GCGAAGCACCTGTGCTGGCTGGGGTTACACACCCCGCAAGAGCACACAGAGAGAGGCC 1341
Db 361 ArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTTCTCTG-----CCCATCGAGCTGCACCCCGACAAGTGCAGCTGCAGCCCATCGAG 1395
Db 381 ProPheLeuTrpMetGlyTyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1396 CTGCGCGCAGAGAGGAGCTGGACCGCTGAACACGACATCCAGAAAGCTGGTGGGCAAGCTGAAC 1455
Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1456 TGGCGCAGCCAGATCTACCCGGGCGATCAAGGTGGCGCAGCTGTGCAAGCTGCTGCGCGC 1515
Db 421 TrpAlaSerGlnIleTyProGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440
QY 1516 GCCAAGGCGCTGACCGACATCGTGCCTGACCGAGGAGCGGAGCTGCAGCTGCAGTGCAGC 1575
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460
QY 1576 AACCGCAGAGATCTTGCAGCGAGCCCGTGCACCGCGCTGTACTACGACCCCGCAGCAAGGACCTG 1635
```



```
Db 281 AlaTyrPheSerValProLeuAspGluAspPheAArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACAGAGACCCCGGATCCGCTACCAAGTACAAAGTCTGCTGCCCGAGGCTGG 1131
Db 301 SerIleAenAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGCGAGCCCGAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerMetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCCGAGATCGTATCTACAG-----GCCCGCTGTAGTGGGAGCCACCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspLeuTyrValGlySerAspLeu 360
QY 1246 GAGATCGGCGAGCACCAGCGCCCAAGATCGAGAGCTGGCAGACACTGCTGGCTGGGCG 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
QY 1306 TTCACCCCGCCGACAAGACACAGAGAGGAGCCCGCTTCCTG-----CCCATCGAG 1359
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGACC 1419
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1420 GTGAACGACATCCAGAGCTGTGGGCAAGCTGAAGCTGGCGCAGCAGCATCTACCCCGCC 1479
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
QY 1480 ATCAGGTGGCGCAGCTGTGCAAGCTGTGGCGGCGCCAGGCGCTGACCGACATCTGTG 1539
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1540 CCCCTGACCGAGAGCGCGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGCAGGCC 1599
Db 461 ProLeuThrGluGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLysGluPro 480
QY 1600 GTGCACGGCGTGTACTAGACCCCGAGCAAGACCTGTGGCGCGAGATCCAGAGCAGGCG 1659
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1660 CACGACCACTGACCTACAGATCTACAGAGACCTTCAAGACCTCAAGACCTGACCGCAG 1719
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnIleTyrGlnIleProPheLysAsnLeuLysThrGlyLys 520
QY 1720 TACGCCAAGATCGGCACCGCCACACCAACGACGTGAAGCAGCTGACCGAGCGCTGCAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1780 AAGATCGCCATGGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTyrGlyLysThrProLysPheLysLeuProIle 560
QY 1840 CAGAGGAGACCTGGAGACCTGTGGACCGACTACTGCGAGCCACCTGGATCCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1900 TGGAGATTCTGTGAACACACCCCGCTGTGTGAAGCTGTGTGATCCAGCTGGAGAAGAGCGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTyrTrpTyrGlnLeuLysGluPro 600
QY 1960 ATCATCGGCGCGAGACCTTCTACTGTGACGCGCGCGCCCAACCGCGAGACCAAGATCGCG 2019
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCGCGCTAGCTACCGACCGCGCGGCGGAGAGATCTGACGCTGACCGAGCCACCC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACCAAGAGACCGAGCTGCAGGCCCATCCAGCTGGCGCTGCAGGACAGCGCGCAGGAGTG 2139
```

```
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCTGTGACCGACAGCCAGTACGCTCGGGCATCATCCAGGCCCGACGACAGAGCC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGACGAGCTGTGTGAACACGAGATCATCGAGCAGCTGTCAAGAGGAGGAGAAGGTGTACCTG 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCGCCGCCAACAGGGCATCGCGCAACGAGCAGATCGACAAGCTGTGTAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGlnGlnValAspLysLeuValSer 720
QY 2320 AAGGCGATCCCGACAGGTGCTCTTCTGACCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 5
T09440
pol polyprotein - human immunodeficiency virus type 1 (strain JRF1) (fragment)
C:Species: human immunodeficiency virus type 1, HIV-1
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T09440
R:Pang, S.; Vinters, H.V.; Akashi, T.; O'Brien, W.A.; Chen, I.S.; Koyanagi, Y.; Namazie,
submitted to the EMBL Data Library, July 1996
A:Reference number: Z16673
A:Accession: T09440
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1003 <PAN>
A:Cross-references: UNIPROT:Q75755; EMBL:U63632; NID:g1465777; PID:g1465779
C:Genetics:
A:Gene: pol
C:Superfamily: pol polyprotein

Alignment Scores:
Pred. No.: 1,01e-151 Length: 1003
Score: 3410.00 Matches: 642
Percent Similarity: 94.45% Conservative: 39
Best Local Similarity: 89.04% Mismatches: 32
Query Match: 74.86% Indels: 8
DB: 2 Gaps: 4

US-09-610-313B-32 (1-2457) x T09440 (1-1003)
QY 220 TTCTTCGCGAGGACCTGGCTTCCCCAGGGCAAGGCCGCGAGTTCCCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACCGCGCAACAGCCCGCCAGCGCGCGAGCTGCAGGTG-----CGGGCGACACCCC 333
Db 21 ThrArgAlaAsnSerProThrArgLysGluLeuGlnValTrpGlyArgAspSerAsnSer 40
QY 334 CGCAGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyAsnValSerPheAsnPheProGlnIleThr 60
QY 388 CTGTGGCAGCGCCCGCTGTGATCAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
QY 448 GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
QY 508 AAGATGATCGCGCGCATCGCGCGCTTCATCAAGGTGGCGCGCGCGCGCGCGCGCGCGCGCG 567
Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeu 120
QY 568 GAGATCTCGCGCAAGAAGCCCATCGGCACTGTGTGATCGGCGCGCGCGCGCGCGCGCGCG 627
Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140
```

```
QY 628 ATCGCCCGCAACATCTGACCCAGCTGGGTGACCCCTGAACTTCCCATCAGCCCCATC 687
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 IleGlyArgAsnLeuThrGlnIleGlyCysThrLeuAsnPheProIle 160
QY 688 GAGACCGTCCCGTGAAGCTGAAGCCCGGATGACCGCCCAAGGTGAACAGTGGGCC 747
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
161 GluThrValProValIleuLysProGlyMetAspGlyProLysValIleGlnTrpPro 180
QY 748 CTGACCGAGGAGAAGATCAAGCCCTGACCCGCTCTCGAGGAGATGGAGAAGAGGCC 807
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200
QY 808 AAGATCACCAGATCGGCCCGGAGAACCCCTACAAACCCCGTGTTCGTCATCAAGAAG 867
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220
QY 868 AAGGACAGCACCAAGTGGCGCAAGCTGGTGACCTTCGCGAGCTGACAAAGCGCACCCAG 927
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysLysThrGln 240
QY 928 GACTTCTGGGAGGTGACCTGGCATCCCGCATCCCGCGCGCTGAAGAAAGAAGAGC 987
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysArgLysSer 260
QY 988 GTGACCGTGTGGACGTGGCGACGCCCTTACAGCGTGCCTCGACGAGACTTCGCG 1047
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280
QY 1048 AAGTACACCGCTTCACCATCCCGAGCATCAACACGAGACCCCGGATCCGCTACGAG 1107
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300
QY 1108 TACAACGTGTGCGCCCGAGGGCTGAAGGGCAGCCCGACATCTTCCAGAGCAGATGACC 1167
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320
QY 1168 AAGATCTCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATCTACCGAG- 1221
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
321 LysIleLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340
QY 1222 CCCCTGTACGTGGCAGGACCTGGAGATCGGCAGACCCCGCGCAAGATCGAGAGCTG 1281
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeu 360
QY 1282 CGCAAGCACCTGTGCTGGGGCTTCACACCCCGCAGCAAGAGACCAAGAGAGGCC 1341
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380
QY 1342 CCCTTCTCTG-----CCCATCGAGCTGCACCCCGCAAGTGGACCGTGCAGCCCATCGAG 1395
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 ProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
QY 1396 CTGCCCCAGAGAGGAGCTGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1455
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
QY 1456 TGGGCCAGCAGATCTACCCCGCATCAAGTGGCGCAGCTGTCAAGCTGCTCGCGCGC 1515
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
QY 1516 GCCAAGCCCTTGACCGCATCGTCCCTCTGACCCGAGGAGCGCGAGCTGGAGCTGGCCGAG 1575
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460
QY 1576 AACCGCAGATCTCGCGGAGCCCGTGCAGCGCTGTACTACGACCCCGCAGCAGAGCTG 1635
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480
QY 1636 GTGCCCCAGATCCAGAGCAGGGCCAGCAGCAGTGGACCTACCGATCTACCGAGGCC 1695
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 IleAlaGluLeuGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500
```

```
QY 1696 TTCAGAACTCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCACACCAACGACGTG 1755
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
501 PheLysIleLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520
QY 1756 AAGCAGCTGACCGAGGCGGTGACAGATCGCATGGAGAGCATCGTCTGCGGGCAG 1815
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
QY 1816 ACCCCCAAGTTCCCGCTGCCCATCCAGAAAGAGACCTGGGAGACCTGTCGACCCGACTAC 1875
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
QY 1876 TGGCAGCGCACCTCGATCCCGAGTGGAGTTCGTGAACACCCCGCTCGTGAAGCTG 1935
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
QY 1936 TGGTACCAGCTGGAGAAGGAGCCCATCATCGCGCGCGAGACCTTCTAGCTGAGCGGCC 1995
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
QY 1996 GCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGACCGGGCGCGCAGAAG 2055
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnArgGlyArgGlnLys 620
QY 2056 ATCTGTAGCTGACCGAGACCAACACGAGACCGAGCTGCAGGCCATCCAGCTGCC 2115
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
QY 2116 CTGCAGACAGCGCAGGAGGTGAACATCGTGTACCGACGACCGATCGCCCTGGGCAATC 2175
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
QY 2176 ATCCAGCGCCAGCCGCAAGAGCGAGCGAGCTGTGTGAACACGATCATCGAGCAGCTG 2235
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
QY 2236 ATCAAGAGGAGAGGTGTATCTGAGCTGGTGGTCCCGCCCAAGAGGCGATCGCGGCAAC 2295
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyLysAsn 700
QY 2296 GAGCAGATCCACAAGCTGTGTGAGCAAGGGCATCGCAGGTGTCTCTCGACGGCATC 2355
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 GluGlnValAspLysLeuValSerAlaGlyIleArgLysValLeuPheLeuAspGlyIle 720
QY 2356 GAT 2358
Db |||
721 Asp 721
```

## RESULT 6

B44001

HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (strain YU-2)  
N;Contains: endonuclease (EC 3.1.-.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly-  
C;Species: human immunodeficiency virus type 1, HIV-1

A;Note: host Homo sapiens (man)  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004

C;Accession: B44001

R;Li, Y.; Hui, H.; Burgess, C.J.; Price, R.W.; Sharp, P.M.; Hahn, B.H.; Shaw, G.M.

J. Virol. 66, 6587-6600, 1992

A;Title: Complete nucleotide sequence, genome organization, and biological properties of  
A;Reference number: A44001; MUID:93021387; PMID:1404605

A;Accession: B44001

A;Molecule type: DNA

A;Residues: 1-1003 <LIV>

A;Cross-references: UNIPROT:P35963; GB:M93258

C;Comment: This protein is synthesized as a gag-pol polyprotein.

C;Genetics: Specific enzymatic cleavages may yield mature proteins including proteinase,

A;Gene: pol

C;Superfamily: pol polyprotein

C;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo

F;S7-155/Product: retropepsin #status predicted <RTP>

F;81/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:

Pred. No.: 1 39e-151 Length: 1003  
 Score: 3407.00 Matches: 643  
 Percent Similarity: 94.31% Conservative: 37  
 Best Local Similarity: 89.18% Mismatches: 33  
 Query Match: 74.80% Indels: 8  
 DB: 1 Gaps: 4

US-09-610-313B-32 (1-2457) x B44001 (1-1003)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGGCGCGAGTTCCTCCCGAGGAGCAG 279  
 Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20  
 QY 280 AACCGCGCCACAGCCCCACAGCCGCGAGCTGCAGGTG-----CGCGGCGACAAACCCC 333  
 Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnAsnSer 40  
 QY 334 CGCAGCGAGGCGCGCGCGAGCGCGAGGCGACCCCTG-----AACTTCCCGCAGATCAC 387  
 Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheSerPheProGlnIleThr 60  
 QY 388 CTGTGGCAGCGCCCTGTGATGACATCAAGTGGCGCGCAGATCAAGAGGCGCCCTGTCTG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCGCGAGCACACCGTGTGGAGGAGATGAGCTGCCCGGCAAGTGGAGGCC 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluGluMetAsnLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCGCGGCGATCGCGCGCTTCATCAAGTGGCGCGCAGTACGACAGATCTGTATC 567  
 Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgIleThrValSerPheSerPheProGlnIleThr 120  
 QY 568 GAGATCTGGCGCAAGAAGCCATCGGCACCGCTGTGATCGCGCGCCCGCCCGTGAACATC 627  
 Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGCTGGCTGACCCCTGAACTTCCCATCGACCCCATC 687  
 Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACCGTGGCGTGAAGCTGAAGCCCGCATGAGCGCGCCCAAGTGAAGCAGTGGCGCC 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGGAGGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAGGAGGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCACCAAGATCGCGCGCGAGAACCCCTCAACACCCCGTGTTCGCCATCAAGAAG 867  
 Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGGACAGCACCAGTGGCGCAAGCTGGTGACTTCCGCGAGCTCAACAGGCGCACCCAG 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGln 240  
 QY 928 GACTTCTGGAGGTGAGTGGGCGATCCCCACCCCGCGCGCTGAAGAAGAAGAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGGAGTGGGCGAGCGCTTACTTTCAGGTGGCCCTTGAGCAGAGACTTCGCG 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuHisGluAspPheArg 280  
 QY 1048 AAGTACACCGCTTCACCATCCCCAGCATCAACACGAGACCCCGCGCATCGCTTACCGAG 1107  
 Db 281 LysTrpThrAlaPheThrIleProSerIleAsnAsnGlnThrProGlyThrArgTrpGln 300  
 QY 1108 TACAACGTGCTGCTCCCGAGGCGTGGAGGGCAGCCCGCAGCATCTTCCAGAGCAGCATGACC 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320

QY 1168 AAGATCTCTGGAGCCCTTCCGCGCGCAACCCCGAGATCGTGATCTACCAG-----GCC 1221  
 Db 321 ThrIleLeuGluProPheArgLysGlnAsnProAspLeuValIleTyrGlnTyrMetAsp 340  
 QY 1222 CCCCTGTACGTGGGCGAGCAGCTGGAGATCGCGCAGCACCGCGCAAGATCGAGGAGCTG 1281  
 Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGCACCTGTCTGGCTGGGGCTTACCAACCCCGCAAGAGCACCAGAGGAGGCC 1341  
 Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCTCTG-----CCCATCGAGTGCACCCCGACCAAGTGCAGCTGCAGCCCATCGAG 1395  
 Db 391 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400  
 QY 1396 CTGCCCGAAGAGGAGCTGGACCGTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAC 1455  
 Db 401 LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420  
 QY 1456 TGGGCGCAGCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGCAAGCTGCTGCGCGCG 1515  
 Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValArgGlnLeuCysLysLeuLeuArgGly 440  
 QY 1516 GCCAAGGCCCTGACGACATCGTCCCTGACCGAGGAGCGCGAGCTGCAGTGGCCGCGAG 1575  
 Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuGluLeuAlaGlu 460  
 QY 1576 AACCGCGAGATCTCGCGAGCGCGTGCACCGCGCTGTACTACGACCCCGCAAGAGCACCTG 1635  
 Db 461 AsnArgGluIleLeuLysGluProValHisGlyValTyrTyrAspProSerLysAspLeu 480  
 QY 1636 GTGGCGCAGATCCAGAGCAGGCGCACACAGTGGACCTACCATCTACGAGGAGGCC 1695  
 Db 481 IleAlaGluIleGlnLysGlnGlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluPro 500  
 QY 1696 TTCGAAGACCTGGAAGCCGCAAGTACGCCAAGATGCGCACCGCCACCAACCAACGACCTG 1755  
 Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaArgThrArgGlyAlaHisThrAsnAspVal 520  
 QY 1756 AAGCAGCTGACCGAGGCGCTGCAGAAGATCGCCATCGAGAGCATCGTGATCTGGGCGAAG 1815  
 Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIleValIleTrpGlyLys 540  
 QY 1816 ACCCCCAAGTTCGCGCTCCCATCCAGAGGAGACCTGGAGACCTGTGAGCCACTAC 1875  
 Db 541 ThrProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpThrGluTyr 560  
 QY 1876 TGGCAGGCGCCTCGATCCCGAGTGGGAGTTTCGTAACACCCCGCTGGTGAAGCTG 1935  
 Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580  
 QY 1936 TGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCAGAGCTTCTACGTGGACGCGCGC 1995  
 Db 581 TrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPheTyrValAspGlyAla 600  
 QY 1996 GCCAACCCGCGAGACCAAGATCGGCAAGCCCGCTACGTAGCAGCCGCGGCGCGGAGAG 2055  
 Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsnLysGlyArgGlnLys 620  
 QY 2056 ATCGTGAGCGCTGACCGAGACCCACCAAGAGACCGAGCTGCAGGCGCATCCAGCTGGCC 2115  
 Db 621 ValValSerLeuThrAspThrAsnGlnLysThrGluLeuGlnAlaIleTyrLeuAla 640  
 QY 2116 CTGAGGACAGCGGCGAGGTGAACATCGTGACCGCAGCAGCAGTACGCTGGGCATC 2175  
 Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
 QY 2176 ATCCAGGCGCCCGCAGCGAGCGAGCGAGCGGCTGGTGAACAGCATCATCGAGGAGCTG 2235  
 Db 661 IleGlnAlaGlnProAspArgSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680

QY 2236 ATCAAGAGGAGAGGTGTACCTGAGCTGGTGGTCCCGCCCAAGAGGAGTCCGGCGCAAC 2295  
 Db 681 ILeIySlySgIuLySValTyLeuAlaTrpValProAlaHISlySgIyIleGlyGlyAan 700  
 QY 2296 GAGCAGATCGACAAGCTGGTGAGCAAGCGCATCCGCAAGGTGCTGTCCTGGAGCGGCATC 2355  
 Db 701 GluInValaPlySLeuValSerAlaGlyIleArgLySValLeuPheLeuAAspGlyIle 720  
 QY 2356 GAT 2358  
 Db 721 Asp 721  
 RESULT 7  
 GNNWA2  
 HIV-1 retropepsin (EC 3.4.23.16) - human immunodeficiency virus type 1 (isolate ARV-2)  
 N;Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
 C;Species: human immunodeficiency virus type 1, HIV-1  
 A;Note: host Homo sapiens (man)  
 C;Date: 17-May-1995 #sequence\_revision 17-May-1995 #text\_change 09-Jul-2004  
 R;Sanchez-Pescador, R.; Power, M.D.; Barr, P.J.; Steimer, K.S.; Stempien, M.M.; Brown-SH  
 Science 227, 484-492, 1985  
 A;Title: Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2).  
 A;Reference number: A04003; MUID:85090453; PMID:2578227  
 A;Accession: A03968  
 A;Molecule type: DNA  
 A;Residues: 1-1003 <S>  
 A;Cross-references: UNIPROT:P03369; GB:K02007; NID:g328658; PIDN:AB59876.1; PID:g328662  
 C;Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
 C;Genetics:  
 A;Gene: pol  
 C;Superfamily: pol polyprotein  
 A;Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucle  
 F;57-147/Product: retropepsin #status predicted <RTP>  
 F;81/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
 Pred. No.: 1.55e-151 Length: 1003  
 Score: 3406.00 Matches: 642  
 Percent Similarity: 94.31% Conservative: 38  
 Best Local Similarity: 89.04% Mismatches: 33  
 Query Match: 74.77% Indels: 8  
 DB: 1 Gaps: 4

US-09-610-313B-32 (1-2457) x GNNWA2 (1-1003)

QY 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGCCCGCGAGTTCGCCAGGAGCAG 279  
 Db 1 PhePheArgGluAAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20  
 QY 280 AACCGCGCCCAACAGCCCGCCAGCGCGAGCTGCAGGTGCGCGC-----GACAAACCC 333  
 Db 21 ThrArgAlaAsnSerProThrArgArgGluLeuGlnValTrpGlyGlyGluAAsnSer 40  
 QY 334 CGCAGCGAGCGCGCGCGAGCGCCAGCGCACCCCTG-----AATCTCCCGAGATCAC 387  
 Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
 QY 388 CTGTGGCAGCGCCCTGCTGGAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCCCTGCTG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleArgIleGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCGCGAGCACCGCTGCGAGAGATGAGCTGCGCGCAAGTGGAGCC 507  
 Db 81 AspThrGlyAlaAspThrValLeuGluGluMetAsnLeuProGlyLysTrpLysPro 100  
 QY 508 AAGATGATCGCGCGCATCGCGGCTTCATCAGGTGCGCGCATCGACGACGATCCTGATC 567  
 Db 101 LysMetIleGlyIleGlyGlyPheIleLysValArgGlnTrpAspGlnIleProVal 120  
 QY 568 GAGATCTGCGCGAAGAGCGCATCGGCACCGTGTGATCGGCGCCACCCCGCTGAGCATC 627  
 Db 121 GluIleCysGlyHisLysAlaIleGlyThrValLeuValGlyProThrProValAsnIle 140

QY 628 ATCGCGCGCAACATGCTGACCCAGCTGGGTGCAACCTGAACTTCCCATCAGCCCCATC 687  
 Db 141 ILeGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAAsnPheProIleSerProIle 160  
 QY 688 GAGACCTGCGCGTGAAGCTGAAGCGCGGATGAGCGGCCCAAGGTGAAGCAGTGGCCC 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGAGGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGGAGAGGAGGC 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCAACAGATCGCCCGCGAGAACCCCTTACAAACACCCCGCTGTTCCCATCAAGAAG 867  
 Db 201 LysIleSerLysIleGlyProGluAAsnProTyAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGGACAGCAACCAAGTGGCGCAAGCTGGTGGATTCCTCGCGAGCTGAACAAGCGCACCC 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAAsnLysArgThrGln 240  
 QY 928 GACTTCTGGGAGTGCAGCTGGGCATCCCGCACCCCGCGCGCTGAAGAAGAAGAAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGAGCGTGGCGCGACCCCTACTTTCAGCGTGGCCCTGCGAGAGTTCGCG 1047  
 Db 261 ValThrValLeuAAspValGlyAspAlaTyPheSerValProLeuAAspLysAspPheArg 280  
 QY 1048 AAGTACACCGCTTCACCATCCCGAGCATCAACAGAGACCCCGCGGATCCGCTACCG 1107  
 Db 281 LysTyThrAlaPheThrIleProSerIleAAsnGlnThrProGlyIleArgTyGln 300  
 QY 1108 TACAAGTGTGTCGCCAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACC 1167  
 Db 301 TyAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCCTGAGCGCTTCCCGCGCCGCAACCCCGAGATCGTGATCTACACG-----GCC 1221  
 Db 321 LysIleLeuGluProPheArgLysGlnAAsnProAspIleValIleTyGlnTyMetAsp 340  
 QY 1222 CCCTGTACGTGGCGAGCGACCTGGAGATCGCGCAGCACCGCGCAAGATCGAGAGCTG 1281  
 Db 341 AspLeuTyValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGCATCTGCTGGCTGGGCTTCACCAACCCCGAGAGCAGACCGAGAGAGAGCC 1341  
 Db 361 ArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCTCTG-----CCCATCGAGCTGCACCCGACAAGTGGACCGTGCAGCCCATCGAG 1395  
 Db 381 ProPheLeuTrpMetGlyTyThrGluLeuHisProAspLysTrpThrValGlnProIleMet 400  
 QY 1396 CTGCCCGAGAGAGAGAGTGCAGCGTGAACGATCAGATCAGAGCTGTGGGCAAGCTGAAC 1455  
 Db 401 LeuProGluLysAspSerTrpThrValAAsnAspIleGlnLysLeuValGlyLysLeuAAsn 420  
 QY 1456 TGGCGCAGCCAGATCTACCCCGCGCATCAAGTGGCGCGAGCTGTCAAGCTGTGCGCGGC 1515  
 Db 421 TrpAlaSerGlnIleTyAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440  
 QY 1516 GCCAAAGCCCTGACCGACATCTGCGCCCTCAGCGAGGAGCGCGAGCTGGAGCTGGCGAG 1575  
 Db 441 ThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGluLeuLeuAlaGlu 460  
 QY 1576 AACCGGAGATCTTGGCGAGCGCGTGCACGGGTGCTACGACCCCGAGCAAGAGACTG 1635  
 Db 461 AsnArgGluIleLeuLysGluProValHisGluValTyTrpAspProSerLysAspLeu 480  
 QY 1636 GTGGCGAGATCCAGAGAGGCGCGAGCGACCTGAGCTTACCATCTACCGAGGAGCC 1695  
 Db 481 ValAlaGluIleGlnLysGlnGlyGlnGlyTrpThrTyGlnIleTyGlnGluPro 500





Db 324 PheGlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleVal 343  
QY 1210 ATCTACCAAG-----GCCCCCTGTGTACCTGGGACGACCTGGAGATCGGCACGACCGC 1263  
Db 344 IleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArg 363  
QY 1264 GCCAAGATCGAGGAGCTGGCAAGCACTGTGCTGGCTGGGCTTCCACACCCCGACAAG 1323  
Db 364 ThrLysIleGluGluLeuArgGlnHisLeuLeuArgTyrGlyLeuThrThrProAspLys 383  
QY 1324 AAGCACCAGAGAGCCCTCTCTG-----CCATCGAGCTGCACCCCGACAGTGG 1377  
Db 384 LysHisGlnLysGluProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyr 403  
QY 1378 ACCGTGACGCCATCGAGCTGCCGAGAGAGCTGCACCGTGAACGACATCCAGAAAG 1437  
Db 404 ThrValGlnProIleValLeuProGluLysAspSerTyrThrValAsnAspIleGlnLys 423  
QY 1438 CTGTGTGGCAAGCTGAACCTGGGCCAGCCAGATCTACCCCGCATCAAGTGTGGCAGCTG 1497  
Db 424 LeuValGlyLysLeuAsnTyrAlaSerGlnIleTyrProGlyIleLysValArgGlnLeu 443  
QY 1498 TGCAAGCTGTGGCGGCCCAAGCCCTGACCGACATCTGCTGCCCTACCGAGAGGCC 1557  
Db 444 CysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAla 463  
QY 1558 GAGCTGGAGCTGGCGGAGAACCGGAGATCTGTGGCGAGCCGTGCAGCGGTGTACTAC 1617  
Db 464 GluLeuGluLeuAlaGluAsnArgGluLeuLysGluProValHisGlyValTyrTyr 483  
QY 1618 GACCCCAAGAGACCTGTGTGGCGGAGATCCAGAACGAGGCGCCACGACCTGACCTAC 1677  
Db 484 AspProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTyrThrTyr 503  
QY 1678 CAGATCTACGAGGACCTTCAAGAACCTGAAGACCGGCAAGTACCGCAATCGGCAC 1737  
Db 504 GlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGly 523  
QY 1738 GCCCACCAACGAGCTGAACGAGCTGACCGAGCGCGTGCAGAGAAGATCGCCATCGAGAGC 1797  
Db 524 AlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSer 543  
QY 1798 ATCTGTATCTGGGCAAGACCCCAAGTTCGCTGCCCTCCATCCAGAAAGGAGACCTGGAG 1857  
Db 544 IleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThrTyrGlu 563  
QY 1858 ACCTGTGGACCGACTACTGCGAGCCACCTGGATCCCGAGTGGAGTTCGTGAACAC 1917  
Db 564 ThrTyrTyrThrGluTyrTyrGlnAlaThrTyrIleProGluTyrGluPheValAsnThr 583  
QY 1918 CCCCCCTGTGTGAAGCTGTGTGTACCGCTGAGAGGAGGCCCATCATCGGCGCGAGACC 1977  
Db 584 ProProLeuValLysLeuTyrTyrGlnLeuGlnLysGluProIleValGlyAlaGluThr 603  
QY 1978 TTCTACGTGGACGGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGTACGTGACC 2037  
Db 604 PheTyrValAspGlyAlaAlaAsnArgGluThrArgLeuGlyLysAlaGlyTyrValThr 623  
QY 2038 GACCGGCGCCCGGCAAGATCGTGAGCTGACCGAGACCAACCAACCAAGAACCGAGCTG 2097  
Db 624 AsnLysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeu 643  
QY 2098 CAGCGCATCCAGCTGGCCCTGAGGACGCGGAGCGGAGGTGACATCGTACCGACGAC 2157  
Db 644 GlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSer 663  
QY 2158 CAGTACGCGCTGGGCATCATCAGCGCCAGCCCAAGAGAGCGAGCGAGCTGTGTGAAC 2217  
Db 664 GlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeuValAsn 683  
QY 2218 CAGATCATCGAGCTGTATCAAGAGAGAGTGTACTCTGAGCTGGTCCCGCCAC 2277

Db 684 GlnIleIleGluGlnLeuIleLysGluLysValTyrLeuAlaTyrValProAlaHis 703  
QY 2278 AAGGCATCGCGCGCAACGACGATCGACAGCTGTGTGAGCAAGGCGATCCGCAAGGTG 2337  
Db 704 LysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysIle 723  
QY 2338 CTGTTCTCGGACCGCATCGAT 2358  
Db 724 LeuPheLeuAspGlyIleAsp 730  
RESULT 9  
GNLJSI  
HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus SIVcpz  
N:Contains: endonuclease (EC 3.1.1.-); retropepsin (EC 3.4.23.16); RNA-directed DNA poly  
C:Species: simian immunodeficiency virus SIVcpz  
A:Note: host Pan troglodytes (chimpanzee)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
C:Accession: S09984  
R:Huet, T.; Cheynier, R.; Meyerhans, A.; Roelants, G.; Wain-Hobson, S.  
Nature 345, 356-359, 1990  
A:Title: Genetic organization of a chimpanzee lentivirus related to HIV-1.  
A:Reference number: S09983; MUID:90259077; PMID:2188136  
A:Accession: S09984  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1027 <HUE>  
A:Cross-references: UNIPROT:P17283; EMBL:X52154  
F:81-180/Product: retropepsin #status predicted <RTP>  
C:Comment: Specific enzymatic cleavages may yield mature proteins including protease, re  
F:105/Active site: Asp (shared with dimeric partner) #status predicted  
A:Gene: pol  
C:Superfamily: pol polypeptide  
C:Keywords: AIDS; aspartic proteinase; endonuclease; hydrolase; immunodeficiency; nucleo  
P:81-180/Product: retropepsin #status predicted <RTP>  
F:105/Active site: Asp (shared with dimeric partner) #status predicted

Alignment Scores:  
Pred. No.: 1,93e-140 Length: 1027  
Score: 3168.00 Matches: 591  
Percent Similarity: 88.38% Conservative: 63  
Best Local Similarity: 79.86% Mismatches: 79  
Query Match: 69.55% Indels: 7  
DB: 1 Gaps: 3

US-09-610-313B-32 (1-2457) x GNLJSI (1-1027)

QY 159 CTGCTGGAAGTGGCGGCAAGGAGGCGCCACGATCAAGGACTGCACCGAGCGCCAGGCCAA 218  
Db 7 LeuLeuAlaValTyrAlaArgGlyThrProAsnGluArgLeuHisArgLysThrGlyGlu 26  
QY 219 CTTCTTCGCGAGACCTGCGCTTCCCGGCAAGCGCGAGTTCGCCAGCGAGCA 278  
Db 27 -PhePheArgGluArgLeuAlaPheProGlnArgGluAlaArgGlnLeuCysAlaGlu 46  
QY 279 GAACCGCGCAACAGCCCGACCGCGAGCTGCAGGTGCGCGC-----GACAAACC 332  
Db 46 nAsnArgThrAsnGlyProThrAspArgGluLeuTyrValProGlyGlyArgGluGluPr 66  
QY 333 CCGCAGCGAGCG 392  
Db 66 oGlyGluGluArgGlyArgGluGlnSerIleSerThrAsnLeuProGlnIleThrLeuTr 86  
QY 393 GCAGCG 452  
Db 86 pGlnArgProLeuIleProValLysValGluGlyGlnLeuCysGluAlaLeuLeuAspTh 106  
QY 453 CGCGCGCGCAACACCTGTGTGGAGAGATGAGCTGCCCGGCAAGTGGAAAGCCCAAGAT 512  
Db 106 xGlyAlaAspThrValIleGluArgIleGlnGlnGlyLeuTyrLysProLysMe 126  
QY 513 GATCGCGCGCATCGCGCGCTTCATCAAGGTGCGCGCAGTACGACCGACGATCCCTGATCGAGAT 572  
Db 126 tileGlyGlyIleGlyPheIleLysValLysGlnPheAspAsnValHisIleGluIl 146



```
QY 481 ATGAGCTGCGCGGAGTGAAGCCCAAGATGATCGCGGCATCGCGGCTTCATCAAG 540
Db 1 MetAsnLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLys 20
QY 541 GTGGCCAGTACGACACAGATCTGATCAGATCTGCGCAAGAGAGCCATCGGACCGTG 600
Db 21 ValLysGlnTyrAspGlnIleAlaIleGluLeuCysGlyHisLysAlaIleGlyThrVal 40
QY 601 CTGATCGGCCCCACCCCGTGAACATCATCGCCGCAACATGCTGACCCAGCTGGGCTGC 660
Db 41 LeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnLeuGlyCys 60
QY 661 ACCCTGAATCTCCCATCAGCCCATCGAGACCGTGCCTGTAAGCTGAAGACCCCGGATG 720
Db 61 ThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMet 80
QY 721 GACGGCCCAAGGTGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTCACCGCC 780
Db 81 AspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuIleGlu 100
QY 781 ATCTCGGAGGAGTGAAGAGAGGCAAGATCACCAAGATCGGCCCGGAGAACCCCTTAC 840
Db 101 IleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGluAsnProTyr 120
QY 841 AACACCCCGCTGTCGCCATCAAGAGAGACAGACCAAGTGGCGCAAGCTGGTGAC 900
Db 121 AsnThrProValPheAlaIleLysLysLysAspGlyThrLysTrpArgLysLeuValAsp 140
QY 901 TTCGCGGAGTGAACAGCGCACCCAGGACTCTGGGAGGTGCAGCTGGGCATCCCCAC 960
Db 141 PheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHis 160
QY 961 CCCGCGCGCTTGAAGAGAGAGCGTGACCGTGTGGAGTGGCGAGCGCTACTTTC 1020
Db 161 ProGlyGlyLeuLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPhe 180
QY 1021 AGCTGCCCCGTGGAGGAGTCTCCGCAAGTACACCGCTTCACATCCCCAGCATCAAC 1080
Db 181 SerValProLeuAspLysAspPheArgLysTyrThrAlaPheThrIleProSerIleAsn 200
QY 1081 AACGAGACCCCGCATCCGTCACGACGATCAACGTCGTCGCCCGCGAGGTGAAGGGCAGC 1140
Db 201 AsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySer 220
QY 1141 CCCAGCATCTTCCAGAGCAGCATCAACAGATCTGAGCGCTTCGCGCGCCGCAACCCC 1200
Db 221 ProAlaIlePheGlnAlaSerMetThrLysIleLeuGluProPheArgLysGlnAsnPro 240
QY 1201 GAGATCGTGATCTACCAAG-----GCCCGCTGTACGTGGCAGCGACCTGGAGTCGC 1254
Db 241 GluIleIleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGly 260
QY 1255 CAGCACCGCGCAAGATCGAGAGTGGCGAAGCAGCTGCTGCGCTGGGCTTCACCAACC 1314
Db 261 GlnHisArgThrLysIleGluLeuArgHisLeuLeuArgTrpGlyPheThr 280
QY 1315 CCCGACAAGAGCACCAAGAGCGCCCTTCTCGCCCATC-----GAGCTGCACCCC 1368
Db 281 ProAspLysLysHisGlnLysGluProProPheLeuTrpIleGlyTyrGluLeuHisPro 300
QY 1369 GACAAGTGGACCGTGCACGCCATCAGCTGCGCGAGAGAGAGCTGACCGTGAACGAC 1428
Db 301 AspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAsp 320
QY 1429 ATCCAGAGCTGTGGGCAAGCTGAATGGGCGCAGCGATCTACCCCGGATCAAGGTG 1488
Db 321 IleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLysVal 340
QY 1489 CGCCAGCTGTCAAGTGTCTCGCGGGCGCCCAAGCGCTTACCGCATCTGTCGCCCTGACC 1548
Db 341 ArgGlnLeuCysArgLeuLeuArgGlyAlaLysAlaLeuThrGluValIleProLeuThr 360
QY 1549 GAGGAGGCGGAGCTGGAGCTGGCGGAGAACCGCGAGATCTCTGCGGAGCGCCGTGCACGGC 1608
```

```
Db 361 LysGluAlaGluLeuGluLeuAlaGluAsnArgGluLeuLysThrProValHisGly 380
QY 1609 GTGTACTACGACCCCAAGGACCTGTGTGCCAGATCCAGAGACGAGCGGCACACGACAG 1668
Db 381 ValTyrTyrAspProSerLysAspLeuValAlaGluLeuGlnLysGlnGlyLeuGlyGln 400
QY 1669 TGGACCTACAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAG 1728
Db 401 TrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLys 420
QY 1729 ATGGGACCCGCCACCAACAGCTGAGCAGCTGACCGAGCGGCGGTGCAGAGATCGCC 1788
Db 421 MetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAla 440
QY 1789 ATGGAGAGCATCGTGATCTGGGCAAGACCCCAAGTTCGCCCTGCCCATCCAGAGGAG 1848
Db 441 ThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIleGlnLysGlu 460
QY 1849 ACCTGGAGACCTGTGTGGACCGACTACTTGGCAGCCCACTGGATCCCCGAGTGGAGTTC 1908
Db 461 ThrTrpGluAlaTrpTrpMetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPhe 480
QY 1909 GTGAGACCCCGCCCTGCTGAGCTGTGTACAGCTGGAGAGAGGCCCATCATCGGC 1968
Db 481 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGly 500
QY 1969 GCCGAGACCTTCTACGTGGAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCCGC 2028
Db 501 AlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGly 520
QY 2029 TACGTGACCAACCGGCGCGCGAGAGATCGTGTGACCGCTGACCGAGACCAACCAAG 2088
Db 521 TyrValThrAspArgGlyArgGlnLysValIleSerLeuThrAspThrThrAsnGlnLys 540
QY 2089 ACCGAGCTGACGSCCATCCAGCTGCGCCCTGACAGACAGCGCAGCGAGGTGAACATCGTG 2148
Db 541 ThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleVal 560
QY 2149 ACCGAGACCGCATGACGCGCTGGGCATCATCCAGCGCCAGCCCGCACAGAGCGAGCGAG 2208
Db 561 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 580
QY 2209 CTGTGTAAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGGTGTACCTGAGCTGGGTG 2268
Db 581 LeuValSerGlnIleIleGluHisLeuIleLysLysGluLysValTyrLeuAlaTrpVal 600
QY 2269 CCGCGCCCAAGGCGCATCGCGGCAACGAGCAGATCCAGAGCTGTGTGAGCAAGGCGCATC 2328
Db 601 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerAlaGlyIle 620
QY 2329 CGCAAGCTGTCTTCTCGGACGCGCATCGAT 2358
Db 621 ArgLysValLeuPheLeuAspGlyIleAsp 630
```

## RESULT 11

T01668

pol polyprotein - human immunodeficiency virus type 1

C;Species: human immunodeficiency virus type 1, HIV-1

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C;Accession: T01668

R;Alizon, M.; Wain-Hobson, S.; Gluckman, J.C.; Sonigo, P.

Cell 46, 63-74, 1986

A;Title: Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isol

A;Reference number: Z14389; MUID:86245056; PMID:2424612

A;Accession: T01668

A;Status: translated

A;Molecule type: mRNA

A;Residues: 1-902 &lt;ALI&gt;

A;Cross-references: EMBL:K03456; NID:g60228; PIDN:CAA28012.1; PID:g60230

C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.: 5.37e-135 Length: 902  
 Score: 3051.00 Matches: 570  
 Percent Similarity: 96.45% Conservatives: 28  
 Best Local Similarity: 91.94% Mismatches: 18  
 Query Match: 66.98% Indels: 4  
 DB: 2 Gaps: 2

US-09-610-313B-32 (1-2457) x T01668 (1-902)

QY 511 ATGATCGGGGATCGCGGGCTTCATCAAGGTGCGCGAGTACGACCAAGATCTGATCGAG 570  
 Db 1 MetIleGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGlu 20

QY 571 ATCTCGGCAAGAGCCATCGCGACCGCTGATCGCGCCCAACCCCGTGAACATCATC 630  
 Db 21 IleCysGlyLysLeuAlaIleGlyThrIleLeuValGlyProThrProValAsnIleIle 40

QY 631 GGCCGCAACATCTGACCCAGCTGGGCTGCACCTCGAACTTCCCATCAGCCCAATCGAG 690  
 Db 41 GlyA-GasnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleGlu 60

QY 691 ACCGTGCGGTGAAGCCCGCATGAGCGCGCATGAGCGCCCAAGTGAAGCAGTGGCCCTG 750  
 Db 61 ThrValProValLysLeuLysProGlyMetAspGlyProArgValLysGlnTrpProLeu 80

QY 751 ACCGAGGAGAAGATCAAGCCCTGACCGCATCTCGAGGAGTGGAGAGAGAGGCGCAAG 810  
 Db 81 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysAspMetGlnLysGlnLys 100

QY 811 ATCACAAGATCGGCGCCGAGAACCCCTACAAACACCCCGCTTCGCCATCAAGAAGAAG 870  
 Db 101 IleLeuLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLys 120

QY 871 GACAGCACCAAGTGGCGCAAGCTGTGAGACTTCCGCGAGCTGAAACAAGCGCACCCAGGAC 930  
 Db 121 AspSerThrLysTrpArgLysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAsp 140

QY 931 TTCTGGAGGTACGTGGGATCCCAACCCCGCGCGCTCAAGNAGNAGAGCGCTG 990  
 Db 141 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 160

QY 991 ACCGTGCTGGAGCGTGGCGACCCCTACTTCAGCTGCCCTCGAGAGGACTTCGCGAAG 1050  
 Db 161 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluAspPheArgLys 180

QY 1051 TACACCGCTTCACCATCCCAAGCATCAACAACGAGACCCCGGATCGCTACCAAGTAC 1110  
 Db 181 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyr 200

QY 1111 AACGTGCTGCCAGGGCTGGAGGGCAGCCCGAGATCGTATCTACCAAGCATGACCAAG 1170  
 Db 201 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 220

QY 1171 ATCTCTGAGCCCTTCGCGCCCGCAACCCCGAGATCGTATCTACCAAG-----GCCGCC 1224  
 Db 221 IleLeuGluProPheArgThrLysAsnProGluIleValIleTyrGlnTyrMetAspAsp 240

QY 1225 CTGTAGTGGGAGCGCATCTGAGATCGGCCAGCACCGCGCAAGATCGAGAGTGGCGC 1284  
 Db 241 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeuArg 260

QY 1285 AAGCACCTGCTGGTGGGGCTTCAACCCCGGACAGAACCAAGAGGAGGCCCCC 1344  
 Db 261 GluHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 280

QY 1345 TTCTCTG-----CCCATCGAGTGCACCCCGCAAGTGAACCGTGCAGCCCATCGAGCTG 1398  
 Db 281 PheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeu 300

QY 1399 CCGGAGAAGGAGCTGGACCGTGAACCGACATCCAGAAGCTGGTGGCGCAAGCTGAAGTGG 1458  
 Db 301 ProAspLysGluSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 320

RESULT 12

B47175  
 reverse transcriptase, AZT-sensitive variant - human immunodeficiency virus type 1  
 C:Species: human Immunodeficiency virus type 1, HIV-1  
 C:Date: 21-Jan-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C:Accession: B47175  
 R:Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.  
 Proc.Natl. Acad. Sci. U.S.A. 90, 25-29, 1993  
 A:Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in th  
 A:Reference number: A47175; MUID:93126353; PMID:7678340  
 A:Accession: B47175  
 A>Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A:Molecule type: DNA

QY 1459 GCCAGCAGATCTACCCCGGCATCAAGGTGCGCGAGTGTGTCAAGCTGTGTGCGCGCGCC 1518  
 Db 321 AlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAla 340

QY 1519 AAGCCCTGACCGACATCGTCCCTGACCGAGGAGCGCGAGCTGGAGTGGCCGAGAAC 1578  
 Db 341 LysAlaLeuThrAspIleValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsn 360

QY 1579 GCGAGATCTCTCGCGGAGCCCGTGCACCGCTGTACTACGACCCCGAGCAAGACCTGGTG 1638  
 Db 361 ArgGluIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeuIle 380

QY 1639 GCCAGATCCAGAAGCAGCGGCACACAGCTGGAGCTACACAGATCTACCAAGAGCCCTTC 1698  
 Db 381 AlaGluIleGlnLysGlnGlyGlnTrpThrTyrGlnIleTyrGlnIleGlnGlnTyr 400

QY 1699 AAGAACCTGAAGACCGGCAAGTACGCCAAGATGCGCACCGCCCAACCAACAGCACTGAAG 1758  
 Db 401 LysAsnLeuLysThrGlyLysTyrAlaArgIleLysSerAlaHisThrAsnAspValLys 420

QY 1759 CAGCTGACCGAGCGCGTGCAGAAATCCCATGGAGAGCATCGTATCTGGGCGCAAGACC 1818  
 Db 421 GlnLeuThrGluAlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThr 440

QY 1819 CCCAAGTTCCGCTCGCCATCCAGAAAGAGACCTGGGAGACCTGGTGACCGACTACTGG 1878  
 Db 441 ProLysPheArgLeuProIleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrp 460

QY 1879 CAGCGCACTGGATCCCGAGTGGAGTTCGTGAACACACCCCGCTCGTGAAGCTGTGG 1938  
 Db 461 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 480

QY 1939 TACCACTGGAGAGAGAGCCCATCATCGCGCGCGAGACTTCTAGCTGGACGCGCGCGCC 1998  
 Db 481 TyrGlnLeuGluThrGluProIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 500

QY 1999 AACCGGAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGCGCGCGAGAGATC 2058  
 Db 501 AsnArgGluThrLysLysGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 520

QY 2059 GTGAGCTGACCGGAGACCAACCAAGACCGAGTGCAGGCGCATCCAGCTGCGCCCTG 2118  
 Db 521 ValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeu 540

QY 2119 CAGGACCGCGGAGGAGGTGAACATCGTACCGGACGACCGAGTACGCTGGGCGATCATC 2178  
 Db 541 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 560

QY 2179 CAGGCGCCAGCCGACAGAGCGAGCGAGCTGGTGAACCAAGATCATCGAGCGCTGATC 2238  
 Db 561 GlnAlaGlnProAspLysSerGluSerGluIleValAsnGlnIleIleGlnLeuIle 580

QY 2239 AAGAGGAGAAGTGTACTCTGAGTGGTGGTCCCGCCCAAGGGCATCGGCGGCAACGAG 2298  
 Db 581 GlnLysAspLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyAsnGlu 600

QY 2299 CAGATCGACAGCTGGTGAAGGCGCATCGCAAGGTGCTGTTCTCGACGCGGATCGAT 2358  
 Db 601 GlnValAspLysLeuValSerSerGlyIleArgLysValLeuPheLeuAspGlyIleAsp 620

A;Residues: 1-559 <MOH>  
 A;Cross-references: UNIPROT:Q9PXX1  
 A;Note: sequence extracted from NCBI backbone (NCBIP:122099)  
 C;Superfamily: pol polyprotein

## Alignment Scores:

Pred. No.:	5, 7e-120	Length:	559
Score:	2728.00	Matches:	510
Percent Similarity:	95.35%	Conservative:	23
Best Local Similarity:	91.23%	Mismatches:	22
Query Match:	59.89%	Indels:	4
DB:	2	Gaps:	2

US-09-610-313B-32 (1-2457) x B47175 (1-559)

Qy	673	CCCATGACCCCGGAGACCGTCCCGTGAAGCTGAAGCCCGGCGATGAGCGGCCCAAG	732
Db	1	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	20
Qy	733	GTGAAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCATCTCGAGGAG	792
Db	21	VallysGlnTrpProLeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGlu	40
Qy	793	ATGAGAGAGGAGGACAGATCACCAAGATCGCCCGGAGACCCCTACCAACACCCCGTG	852
Db	41	MetGlyLysGluGlyLysIleSerLysIleGlyProGluAsnProTyrAsnThrProVal	60
Qy	853	TTCCGCCATCAAGAGAGGACAGACCAAGTGGCGGAGCTGGTGGACTTCCCGGAGCTG	912
Db	61	PhelaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	80
Qy	913	AACAAGCGACCCAGGACTTCTGGAGGTGAGCTGGGCGATCCCGGAGCTG	972
Db	81	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	100
Qy	973	AAGAGAGAGAGCGGTGACCGTCTGACGTGGCGGAGCGCTACTTACGGTCCCGCTG	1032
Db	101	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	120
Qy	1033	GACGAGACTTCCCAAGTACACCGCTTACCATCCCGGAGTCAACCAACGAGACCCCG	1092
Db	121	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	140
Qy	1093	GGCATCCGCTACCAAGTACAAGTCTGCCAGGGCTGGAGGCGACCCCGAGCATCTTC	1152
Db	141	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	160
Qy	1153	CAGAGCAGCATGACCAAGATCTGGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC	1212
Db	161	GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspIleValIle	180
Qy	1213	TACCAG-----GCCCGCTGTAGTGGCGAGCGACCTGGAGATCGCGGAGCGCGCC	1266
Db	181	TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr	200
Qy	1267	AAGATCAGGAGCTGCGCAAGACACTGCTGGCTGGGGCTTACACACCCCGGCAACAAG	1326
Db	201	LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyLeuThrThrProAspLysLys	220
Qy	1327	CACAGAGAGGAGGCGCCCTTCCTG-----CCCATCGAGTGCACCCCGCAAGTGAAC	1380
Db	221	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	240
Qy	1381	GTGCAGCCCATCGAGTCCCGGAGAGAGAGCTGGACCTGGACGATCCGACATCCGAGCTG	1440
Db	241	ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu	260
Qy	1441	GTGGCGAAGTGAACCTGGCGCAGCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTC	1500
Db	261	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValArgGlnLeuLys	280
Qy	1501	AAGTGTGCGCGCGCAAGGCGCTGACCGACATCGTCCCTGACCGGAGGCGCGAG	1560

Db	281	LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleProLeuThrGluGluAlaGlu	300
Qy	1561	CTGAGCTGGCGGAGAACCGGAGATCTCGCGGAGCCCGGCGGCTGTACTACGAC	1620
Db	301	LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp	320
Qy	1621	CCCAGCAAGACCTGGTGGCGGAGATCCAGAAAGCAGGCGCCACGACCTGACCTAC	1680
Db	321	ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGln	340
Qy	1681	ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGAGTACGCAAGATCGGACCCG	1740
Db	341	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla	360
Qy	1741	CACACCAACACGCTGAAGCAGCTCACGAGGCGCTGCAGAAAGATCGCCATCGAGAG	1800
Db	361	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle	380
Qy	1801	GTGATCTGGGCAAGACCCCAAGTTCGGCTGCCATCCAGAGGAGACCTGGGAGACC	1860
Db	381	ValIleTrpGlyLysThrProArgPheLysLeuProIleGlnLysGluThrTrpGluThr	400
Qy	1861	TGGTGGACCGACTACTCGCAGGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCC	1920
Db	401	TrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	420
Qy	1921	CCCCTGCTGAAGCTGTGTACCATCGAGAGGAGCCCATCATCGGCGCGAGACCTTC	1980
Db	421	ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPhe	440
Qy	1981	TACGTGACCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGGCTACGTGACCC	2040
Db	441	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	460
Qy	2041	CGGGCGCGGAGAGATCGTGTGAGCTGACCGGAGACCAACCAAGAGACCGAGCTGAG	2100
Db	461	LysGlyArgGlnLysValValProLeuThrAsnThrThrAsnGlnLysThrGluLeuGln	480
Qy	2101	GCATCCAGCTGGCGCTGCGAGGACGCGGAGGAGTGAACATCGTGACCGAGCCAG	2160
Db	481	AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln	500
Qy	2161	TACGCTCGGCGCATCATCCAGGCGCCAGCCGAGAGGAGGAGCGAGCTGGTGAACCG	2220
Db	501	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln	520
Qy	2221	ATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACTGAGCTGGTGGCGCCGACAG	2280
Db	521	IleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys	540
Qy	2281	GGCATCGGCGCAACGAGCAGATCGCAAGCTGTGAGCAAGGCGCATCCGCAAGGTG	2337
Db	541	GlyIleGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleAlaGlyVal	559

## RESULT 13

A47175

reverse transcriptase, AZT-resistant variant - human immunodeficiency virus type 1  
 C;Species: human immunodeficiency virus type 1, HIV-1  
 C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C;Accession: A47175

R;Mohri, H.; Singh, M.K.; Ching, W.T.; Ho, D.D.

Proc. Natl. Acad. Sci. U.S.A. 90, 25-29, 1993

A;Title: Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the

A;Reference number: A47175; PMID:93126353; PMID:7678340

A;Accession: A47175

A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra

A;Molecule type: nucleic acid

A;Residues: 1-559 &lt;MOH&gt;

A;Cross-references: UNIPROT:Q9PXX2

A;Note: sequence extracted from NCBI backbone (NCBIP:122100)

C;Superfamily: pol polyprotein

Alignment Scores:

Pred. No.:	4.88e-119	Length:	559
Score:	2708.00	Matches:	506
Percent Similarity:	95.17%	Conservative:	26
Best Local Similarity:	90.52%	Mismatches:	23
Query Match:	59.45%	Indels:	4
DB:	2	Gaps:	2
US-09-610-313B-32 (1-2457) x A47175 (1-559)			
QY	673	CCCATCAGCCCCATCGAGACCTGCGCGTGAAGCCCGGCGATGAGCGGCCCAAG	732
Db	1	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	20
QY	733	GTGAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCATCTGCGAGGAG	792
Db	21	ValLysGlnTrpProLeuThrGluLysIleLysAlaLeuValGluLysCysThrGlu	40
QY	793	ATGGAGAAGGAGGCAAGATCACCAAGATCGGCCCGCGAGAACCCCTACAAACACCCCGTG	852
Db	41	MetGluLysGluGlyIleSerLysIleGlyProGluAsnProTyrAsnThrProVal	60
QY	853	TTCCGCATCAGAAGAGACAGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG	912
Db	61	PheAlaIleLysLysLysAspSerThrArgTrpArgLysLeuValAspPheArgGluLeu	80
QY	913	AACAGCGCACCCAGGACTTCTGGGAGTGCAGCTGGGCGATCCCGCCCGCGCGCTG	972
Db	81	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	100
QY	973	AAGAAGAAGAAGCGTGCACCGTGTGGACGTGGCGCGACGCTACTTTCAGCGTGGCCCTG	1032
Db	101	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	120
QY	1033	GACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGCATCAACACGAGACCCCG	1092
Db	121	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	140
QY	1093	GCATCCGCTACGATACAACTGCTGCCAGGGCTGGAGGGCAGCCCGAGCATCTTC	1152
Db	141	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	160
QY	1153	CAGACGAGCATACCAAGATCTCGAGCCCTTCCCGCGCCGCAACCCCGAGATCGTGATC	1212
Db	161	GlnSerSerMetThrLysIleLeuGluProPheArgLysGlnAsnProAspMetValIle	180
QY	1213	THACAG-----GCCCGCTGTACGTGGCGACGACCTGGAGATCGCGCAGCACCGCGC	1266
Db	181	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThr	200
QY	1267	AGATCGAGGAGCTGGCAAGCACCTGCTGGCTGGGCTTCACCCCGCCGACAGAGAG	1326
Db	201	LysIleGluGluLeuArgGlnHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	220
QY	1327	CACCAAGAAGGAGCCCGCTTCCTG-----CCCATCGAGCTGCACCCCGCAAGTGGACC	1380
Db	221	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	240
QY	1381	GTGAGCCCATCAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTG	1440
Db	241	ValGlnProIleValLeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeu	260
QY	1441	GTGGCGAGCTGAACCTGGCGCAGCAGATCTACCCCGCATCAAGTGGCGCGCTGTGC	1500
Db	261	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValLysGlnLeuCys	280
QY	1501	AAAGTGTCTCGCGGGGCCAAGGCCCTTACGACATCGTCCCGCTGACCGAGAGGCCGAG	1560
Db	281	LysLeuLeuArgGlyThrLysAlaLeuThrGluValIleGlnLeuThrGluGluAlaGlu	300
QY	1561	CTGAGCTGGCGCAGAACCGCGAGATCTCTGCGCGAGCCCGTGCACGCGGTGTACTACGAC	1620
Db	301	LeuGluLeuAlaGluAsnArgGluIleLeuArgGluProValHisGlyValTyrTrpAsp	320

QY	1621	CCAGACAGGACCTGGTGGCCGAGATCCAGAAAGCAGGGCCACGACCGAGTGGACCTACCA	1680
Db	321	ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyGlnTrpThrTyrGln	340
QY	1681	ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACGCGCAAGATGCGCACCGC	1740
Db	341	IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMetArgGlyAla	360
QY	1741	CACACACAGAGCTGAAGCAGCTGACCGAGCGCTGCGAAGATCCCATGGAGAGCATC	1800
Db	361	HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThrGluSerIle	380
QY	1801	GTGATCTGGCGCAAGACCCCAAGTTCGCTGCCCATCCAGAAAGAGACCTGGGAGACC	1860
Db	381	ValIleTrpGlyLysIleProArgPheLysLeuProIleGlnLysGluThrTrpGluAla	400
QY	1861	TGGTGGACCGACTACTGGCAGCCACCTGGATCCCGAGTGGGAGTTCGTGAACACCCCG	1920
Db	401	TrpTrpIleGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	420
QY	1921	CCCTGGTGAAGCTGGTACCGCTGAGAGGAGGCCCATCATCGCGCGCGAGACCTTC	1980
Db	421	ProLeuValLysLeuTrpTyrGlnLeuGlnLysGluProIleValGlyAlaGluThrPhe	440
QY	1981	TACGTGGACCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACCTGACCCGAC	2040
Db	441	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsn	460
QY	2041	CGGCGCGCGCAGAAGATCGTGAGCTGACCGAGACCCCAACACCAAGACCGAGCTGCAG	2100
Db	461	LysGlyArgGlnLysValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGln	480
QY	2101	GCCATCCAGCTGGCCCTGCAGACAGCGCGAGCGAGTGAACNTCGTCACCGACAGCCAG	2160
Db	481	AlaIleHisLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGln	500
QY	2161	TACGCGCTGGGCATCATCCAGCGCCGACCCAGACAGAGCGAGCGAGTGGTGAACACAG	2220
Db	501	TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGln	520
QY	2221	ATCATCGAGCAGCTGATCAAGAAGAGAGTGTACTCTGAGCTGGGTGCCCGCCACCAAG	2280
Db	521	IleIleGluGluLeuIleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLys	540
QY	2281	GGCATCGCGCCACAGCAGCATCGACAAGCTGGTGGAGAGGGGCATCCGCAAGCTG	2337
Db	541	GlyIleGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArgLysVal	559
RESULT 14			
S46347			
pol polyprotein - simian immunodeficiency virus SIVagm (isolate SAB-1)			
C:Species: simian immunodeficiency virus SIVagm			
A:Variety: isolate SAB-1			
C:Date: 25-Dec-1994 #sequence_revision 14-Feb-1997 #text_change 26-Aug-1999			
C:Accession: S46347			
R:Jin, M.J.; Hui, H.; Robertson, D.L.; Mueller, M.C.; Barre-Sinoussi, F.; Hirsch, V.M.;			
EMBO J. 13, 2935-2947, 1994			
A:Title: Mosaic genome structure of simian immunodeficiency virus from West African green			
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993			
A:Reference number: S46335; MUID:94298785; PMID:8026477			
A:Accession: S46347			
A:Status: nucleic acid sequence not shown; translation not shown			
A:Molecule type: DNA			
A:Residues: 1-1039 <JIN>			
A:Cross-references: EMBL:U04005; NID:9466229; PIDN:AAA21505.1; PID:9466231			
A:Experimental source: isolate SAB-1; sabaeus monkey			
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993			
A:Note: this reading frame extends between two stop codons and does not begin with a sta			
C:Genetics:			
A:Gene: pol			
C:Superfamily: pol polyprotein			
Alignment Scores:	2.61e-100	Length:	1039
Pred. No.:			



Score:	2307.00	Matches:	436
Percent Similarity:	73.54%	Conservative:	120
Best Local Similarity:	57.67%	Mismatches:	154
Query Match:	50.65%	Indels:	46
DB:	2	Gaps:	8
US-09-610-313B-32 (1-2457) x S46347 (1-1039)			
QY	220	TTCTTCCGCGAGGACCTGCGCTTCCCGGAGGCGC-----AAGGCCCGCGAGTTCCTCCAGC	273
DB	1	PhePheArg-----ValTrpProLeuGlyGlnArgGluThrGlnGluPheProSer	17
QY	274	GAGCAGAAACCGCGCAAC-AGCCCCACAGCCCGGAGCTGCGAGTCCGCGGC-----	324
DB	18	AspLeuHisGlnThrAsnSerProAsnGlyThrGlyLeuGlnGlnAlaGlyGlyLys	37
QY	325	-----GACAAACCCCGCAGCGAGGCGC-----	345
DB	38	LeuValCysArgGlnThrSerAspGlnArgThrArgAlaArgArgSerAsnSerPro	57
QY	346	-----GGCGCGCGAGCGC-----	357
DB	58	VallysAlaValCysCysSerGlyGluThrAlaGluThrAlaValAlaLysProLeuAla	77
QY	358	-----CAGGCGACCTGAACCTTCCCGCAGATCACCTGTGGCAGCGGCC	402
DB	78	ThrThrGluProLeuArgGlyGlyLeuGlnLeuProGlnValSerLeuTrpArgArgPro	97
QY	403	CTGGTGACATCAAGTGGCGGCAGATCATCAAGGAGGCCCTGCTGGACACCGCGCGGAC	462
DB	98	MetLysThrValTyrIleGluGlyGlnLysValThrAlaLeuLeuAspThrGlyAlaAsp	117
QY	463	GACACCGTGTGGAGGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATCATCGCGGC	522
DB	118	AspSerValIleGlnGlyIleGluLeuGlyAspAsnTrpLysProArgIleIleGlyGly	137
QY	523	ATCGCGGCTTCATCAAGTGGCGGCAGTACGACCATCTGTGATCGAGATCTGCGGCAAG	582
DB	138	IleGlyGlyCysIleAsnValLysAlaTyrHisAsnGlnGluValLysIleGluAspLys	157
QY	583	AAGCCATCGGACCGTGTGATCGGCGCCACCCCGTGAACATCATCGCGCGCAACATG	642
DB	158	ThrCysLysAlaThrIleLeuValGlyGluThrProValAsnIleIleGlyArgAsnVal	177
QY	643	CTGACCCAGCTGGCTGACCTCACTTCCCATCGACCCCATCGAGCGTCCCGTG	702
DB	178	LeuAlaGlnLeuGlyValThrLeuAsnLeuThrGlnArgGluIleGluProIleLysVal	197
QY	703	AAGTGAAGCCCGGATGAGCGGCCCAAGGTGAAGCATGCGCCCTGACCGAGAGAAG	762
DB	198	HisLeuLysProGlyGlnAspGlyProArgIleArgGlnTrpProLeuSerLysGluLys	217
QY	763	ATCAAGCCCTGACCGCATCTGCGAGAGATGAGAGGAGGCGCAAGATCACCAATC	822
DB	218	IleGluAlaLeuLysAlaIleCysGluAspLeuGlyHisLeuGluArgIle	237
QY	823	GGCCCGGAGAACCTTACACACCCCGTGTCCGCATCAAGAAGAGAGGAGAGGAGGAG	882
DB	238	GlyProGluAsnProTyrAsnThrProValPheAlaIleArgLysLysAspLysThrGln	257
QY	883	TGGCGCAAGCTGTGGACTTCCCGGAGCTGAACAGCGCACCCAGGACTTCTGGGAGGTG	942
DB	258	TrpArgIleLeuMetAspPheArgGlnLeuAsnLysSerThrGlnAspPheGlnGluVal	277
QY	943	CAGTGGGATCCCGCCACCCCGCGCTGAAGAAGAGAGAGCGGTGACCGTCTGGAC	1002
DB	278	GlnLeuGlyIleProHisProAlaGlyLeuGlnGlnArgGlnGlnIleThrValLeuAsp	297
QY	1003	GTGGCGGACGCTACTTTCAGGTGCGCTGACGAGGACTTCCGCAAGTACACCGCTTC	1062
DB	298	IleGlyAspAlaTyrPheSerCysProLeuAspProAspPheGlnLysTyrThrAlaPhe	317
QY	1063	ACCATCCCGAGCATCAACACAGAGACCCCGGCATCCGCTACCATGATCAACGCTGCTGCC	1122

DB	318	ThrIleProSerValAsnAsnArgGluProGlyIleArgTyrGlnTyrLysValLeuPro	337
QY	1123	CAGGCTGGAGGCGCGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTCTGGAGCCC	1182
DB	338	GlnGlyTrpLysGlySerProThrIlePheGlnThrThrAlaAsnLysIleLeuGlnGlu	357
QY	1183	TTCCGCGCGCCGCAACCCGAGATCGTATCTACACAG-----GCCCGCTGTACGTGGCG	1236
DB	358	PheArgGlnLysAsnProAspValAspIleTyrGlnTyrMetAspAspMetLeuIleAla	377
QY	1237	AGCGACCTGGAGATCGCGCAGCAGCGGCCCAAGATCGAGGAGCTGCGCAAGCATCTGTG	1296
DB	378	SerAspArgProLysAlaGluHisLeuValMetValGlnGlnLeuArgAspTyrLeuGlu	397
QY	1297	CGCTGGGCTTCCACCCCGCAAGAACGACAGGAGGAGGCGCCCTCTCTG-----	1350
DB	398	ThrTrpGlyPheLysThrProGluLysLysPheGlnLysAspProProTyrLeuTrpMet	417
QY	1351	CCCATCGAGCTGCAACCCGACAGTGGACCGCTGAGCGCCCATCGAGCTGCCGAGAGGAG	1410
DB	418	GlyTyrGluLeuTyrProLysLysTrpGlnLeuGlnGluIleThrLeuProGluArgGlu	437
QY	1411	AGCTGGACCTGGAACGACATCCAGAACTGTGTGGCAAGCTGAACCTGGCGCAGCATC	1470
DB	438	GluTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle	457
QY	1471	TACCCCGCATCAAGTGGCGCAGCTGTGAAGCTGTGCGCGCGCCCAAGCGCTGACC	1530
DB	458	TyrThrGlyIleLysThrLysHisLeuCysArgLeuIleArgGlyAlaArgProLeuThr	477
QY	1531	GACATCTGCGCCCTGACCGAGGCGGAGCTGAGCTGGCGGAGAACCGGAGATCTCTG	1590
DB	478	GluIleValGlnTrpThrGluGluAlaGluLeuGluGluGluAsnArgGlnIleLeu	497
QY	1591	CGCGAGCCGTGACGCGTGTACTAGACCCCGCAGCAAGGAGCTGTGTGGCGCAGATCCAG	1650
DB	498	ArgGlnLysGlnGlnGlyGlnTyrTyrAspProAlaLeuProLeuArgAlaLysValLeu	517
QY	1651	AAGCAGGCGCCACGACGAGTGGACCTTACCATCTTCCAGGAGCCCTTCAAGACTGAG	1710
DB	518	LysLeuGlyAspGlyGlnTrpGlyTyrGlnIleTyrGlnProGluAsnLysIleLeuLys	537
QY	1711	ACCGGCAAGTACGCAAGATGCGCAGCGCCGACCAACACGACGAGTGAAGCATCTGACCG	1770
DB	538	ValGlyLysTyrAlaLysIleLysThrAlaHisThrAsnGluLeuArgMetLeuAlaGly	557
QY	1771	GCGGTGAGAAAGTGGCCATGCGAGAGCATCTGTGATCTGGGGCAAGACCCCAAGTTCGCG	1830
DB	558	LeuValGlnLysIleGlyLysGluSerIleValIleTrpGlyGlnIleProIleMetGlu	577
QY	1831	CTGCGCATCCAGAGGAGACTGGGAGACCTGGGACCTGTGGACCGGACTACTGTGACCGC	1890
DB	578	LeuProValGluArgGluLeuTrpGluTrpSerAspTyrTrpGlnValThrTrp	597
QY	1891	ATCCCGCAGTGGGAGTTCGTGAACACACCCCGCTGTGTGAAGCTGTGTGACCGCTGAG	1950
DB	598	IleProGluTrpGluMetValSerThrProGlnLeuIleArgLeuTrpTyrLysLeuVal	617
QY	1951	AAGAGCCCATCATCTGCGCGCGAGACCTTCTACGTGACCGCGCGCCCAACCGCGAGACC	2010
DB	618	LysAspProIleProGlyGluAlaValTyrTyrValAspGlyAlaAlaAsnArgAsnSer	637
QY	2011	AAGATCGCGCAGGCGCGCTAGTGACCGCGCGCGCGCGGCGGAGATCTGTGAGCTGACC	2070
DB	638	LysGluGlyLysAlaGlyTyrLeuThrAspArgGlyAspGlnLysValValAlaLeuGlu	657
QY	2071	GAGACCAACCAAGAGACCGAGCTCAGGCCATCCAGCTGGCGCTGGCAGCAGCGCGC	2130
DB	658	AsnThrThrAsnGlnLysAlaGluLeuGluAlaIleLeuLeuAlaLeuArgAspSerGly	677
QY	2131	AGCGAGTGAACATCTGTGACCGCAGCAGCTGCGCTGCGGCTATCTCCAGCGCGCCGCC	2190

Db 678 SerLysValAsnIleIleThrAspSerGlnTyrAlaMetGlyIleAlaGlyGluPro 697  
QY 2191 GACAGAGCGGAGCGAGCTGTGTAACACAGATCATCGACGACTATCAAGAGAGAG 2250  
Db 698 ThrGluSerAspAsnAsnIleValGlnGlnIleIleGluLeuIleGlyLysGluAla 717  
QY 2251 GTGTACCTGAGCTGGTGGCCGCCACAGGGCATCGGGCAAGAGAGAGAGAGAG 2310  
Db 718 ValTyrIleAlaTrpValProIleHisLysGlyValGlyAsnGluIleAspLys 737  
QY 2311 CTGTGAGCAAGGGCATCGCAAGGTGTCTTCCTGAGCGGATCGAT 2358  
Db 738 LeuValSerGlnGlyIleArgGlnValLeuPheLeuAspArgIleGlu 753  
RESULT 15  
SS3092  
pol polyprotein - human immunodeficiency virus type 2  
C:Species: human immunodeficiency virus type 2, HIV-2  
C:Date: 08-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
C:Accession: SS3092  
R:Becker, M.; Zorr, B.; Becker, A.; Habermehl, K.O.  
submitted to the EMBL Data Library, March 1995  
A:Description: Molecular and phylogenetic characterisation of a Guinea-Bissau-derived hu  
A:Reference number: SS3091  
A:Accession: SS3092  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1055 <BEC>  
A:Cross-references: UNIPROT:O73194; EMBL:Z48731  
C:Superfamily: pol polyprotein  
C:Keywords: polyprotein  
Alignment Scores:  
Pred. No.: 6,37e-94 Length: 1055  
Score: 2170.00 Matches: 413  
Percent Similarity: 70.21% Conservative: 129  
Best Local Similarity: 53.50% Mismatches: 187  
Query Match: 47.64% Indels: 43  
Db: 2 Gaps: 11  
US-09-610-313B-32 (1-2457) x SS3092 (1-1055)  
QY 156 GGGCTGCTGGAAGTGGCGGCAAGGAGGCGCCACACAGATGAAGGACTGCAC---CGAGCGCA 212  
Db 3 GlyLeuLeuGluMetTrpGln-----AspArgThrTyrHisGlyLysValPro 18  
QY 213 GCCCAACTT-----CTCCG 227  
Db 19 ArgGlnThrGlyPhePheArgAspTrpProLeuGlyLysGluAlaProGlnLeuPro 38  
QY 228 CGAGGACTGGCTTCCCGGCAAGGCGCGGAGTTCCCGAGGAGAGAA-----281  
Db 39 ArgGlyProGlySerAlaGlyAlaAsnThrAsnSerThrProSerArgSerSerGly 58  
QY 282 CGCGCCCAACACCCACGCGCGCA-----GTCGAGGTGGCGGCGCAACCCCG 335  
Db 59 ProThrGlyGluIleTyrAlaAlaArgGluLysAlaGluAlaGluArgGluThrIle 78  
QY 336 CAGCGAGCGCGCGC-----CGAGCGCA-----GGCACCCTCAACTT-----374  
Db 79 GlnArgGlyAspArgGlyLeuAlaProArgAlaGlyLysAspThrMetGlnGlyAsp 98  
QY 375 -----CCCCAGATCACCTGTGGAGCGCCCGGAGTTCCCGAGGAGAGAA-----416  
Db 99 AsnArgGlyPheAlaAlaPro-GlnPheSerLeuTrpAsnArgProValValThrAlaHi 118  
QY 417 GTGTGGCGCGAGATCAAGAGCGCTGTGTGACACCGCGCGGAGACACCGTGTGGA 476  
Db 118 sIleGluGlnGlnProValGluValLeuLeuAspThrGlyAlaAspSerIleValAl 138  
QY 477 GGAGATGACCTCGCGCGCAAGTGGAAAGCCCAAGATGATCGCGGCGGCTTCAT 536  
Db 138 aglyIleGluLeuGlySerAsnTyr-SerProLysIleValGlyIleGlyPheIle 158

QY 537 CAAGTGGCGCCAGTAGCACGACCATCTGATCGAGATCTGCGCGCAAGAGCCCATCGGCAC 596  
Db 158 eAsnThrLysGlyTyrLysAsnValGluIleGluValLeuGlyLysArgValArgAlaTh 178  
QY 597 CGTGTGATCGCGCCCGCCCGTGAACATCATCGCGCGCAACATGCTGACCCACCTGGG 656  
Db 178 rIleMetThrGlyAspThrProIleAsnIlePheGlyArgAsnIleLeuThrAlaLeuGl 198  
QY 657 CTGCACCTGAATCTTCCCATCAGCCCATCGAGACCGTGCCTGCGTGAAGCTGAAGCCCGG 716  
Db 198 yMetSerLeuAsnLeuProValAlaLysIleGluProIleLysIleMetLeuLysProGl 218  
QY 717 CATGGACGCGCCCAAGGTGAAGCGTGGCCCTGCACCGAGGAGAGATCAAGGCCCTGAC 776  
Db 218 yLysAspGlyProLysLeuArgGlnTrpProLeuThrLysGluLysIleGluAlaLeuL 238  
QY 777 CGCCATCTGCGAGGAGATGAGAGAGGAGGCAAGATCACCAAGATCGCGCCCGGAGAACCC 836  
Db 238 sGluIleCysGluLysMetGluArgGluGlnLeuGluAlaProProThrAsnPr 258  
QY 837 CTACAACACCCCGTGTTCGCATCAAGAGAAGAGACAGCACCAAGTGGCGCAAGCTGT 896  
Db 258 ofYrAsnThrProThrPheAlaIleArgLysLysAspLysAsnLysTrpArgMetLeuI 278  
QY 897 GGACTTCCGCGAGCTGAACAGCGCCACCGAGCTTCTGGGAGGTGCGAGCTGGGCATCCC 956  
Db 278 eAspPheArgGluLeuAsnLysValThrGlnAspPheThrGluIleGlnLeuGlyIlePr 298  
QY 957 CCACCCCGCGCTGAAGAGAAGAGAGCGTGACCGTGTGAGACGTGGGCGAGCGCTTA 1016  
Db 298 oHisProAlaGlyLeuAlaLysLysArgArgIleThrValLeuAspValGlyAspAlaTy 318  
QY 1017 TTTGAGCTGCCCTGAGAGAGACTTCCGCAAGTACACCGCTTTCACATCCCGAGAT 1076  
Db 318 rPheSerIleProLeuHisGluAspPheArgGlnTyrThrAlaPheThrLeuProSerVa 338  
QY 1077 CAACAACGAGACCCCGCATCCGCTACAGTACACAGTGTGCTGCGAGGCGTGAAGGG 1136  
Db 338 lAsnAsnAlaGluProGlyLysArgTyrIleTyrLysValLeuProGlnGlyTrpLysGl 358  
QY 1137 CAGCCCGAGCATCTTCCAGAGAGCATGACCAAGATCTCGGAGCCCTTCCCGCGCCGCAA 1196  
Db 358 ySerProAlaIlePheGlnTyrThrMetArgGlnValLeuGluProPheArgLysAlaAs 378  
QY 1197 CCGCGAGATCGTATC-----TACCAGCCCGCTGTACGTGGCGACGACCTCGAGAT 1250  
Db 378 nGlnAspValIleIleIleGlnTyrMetAspAspIleLeuIleAlaSerAspArgThrAs 398  
QY 1251 CGGCAGCACCGCGCCCAAGATCGAGGAGCTCGCAAGCACCTGTGCTGCGTGGGGCTTCA 1310  
Db 398 pLeuGluHisAspArgValValLeuGlnLysGluLeuLysGluLeuAsnSerLeuGlyPheSe 418  
QY 1311 CACCCCGCACAGAACACCAAGAGGAGCGCCCTTC-----CTGCCCATCGAGCTGCA 1364  
Db 418 rThrProAspGluLysPheGlnLysAspProProTyrArgTrpMetGlyTyrGluLeuTr 438  
QY 1365 CCGCGACAAAGTGGACCGTGTGACCGCCATCGAGTGTCCCGAGAGAGAGAGCTGACCGTGA 1424  
Db 438 pProThrLysTrpLysLeuGlnLysIleGlnLeuProGlnLysGluValTrpThrValAs 458  
QY 1425 CGACATCCAGAGCTGTGGCGAGCTGAATGGCGCCAGCCAGATCTACCCCGGATCAA 1484  
Db 458 nAspIleGlnLysLeuValGlyValLeuAsnTrpAlaAlaGlnIleTyrProGlyIleL 478  
QY 1485 GTGTGGCGAGCTGTCAAGCTGTGCGCGCGCGCAAGGCCCTGACCGGACATCGTCCCT 1544  
Db 478 sThrLysHisLeuCysArgGluIleArgGlyLysMetThrLeuThrGluGluIleGlnTr 498  
QY 1545 GACCGAGGAGCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTCTGCGGAGCCCGTGA 1604  
Db 498 pThrGluLeuAlaGluAlaGluLeuGluGluAsnArgValIleLeuSerGlnGluGlnGl 518

```

QY 1605 CGCGGTGTACTACGACCCCGAGGACCTGGTGGCGGAGATCCAGAGAGCGGCCACGA 1664
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
518 uGlyHisTyrTyrGlnGluGluLeuGluAlaThrValGlnLysAspGlnAspAs 538
QY 1665 CCAGTGGACCTACGACATCTACGAGAGCCCTTCAAGACCTCAAGACCGGCAAGTACGC 1724
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
538 nGlnTrpThrTyrLysValHisGlnGlu---LysThrLeuLysValGlyLysTyrAl 557
QY 1725 CAAGATGCGCACCGCCACACCAACGACGCTGAAGCAGCTGACCGAGGCGCTGCAGAGAT 1784
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
557 aLysValLysAsnThrHisThrAsnGlyValArgLeuLeuAlaGlnValGlnLysIl 577
QY 1785 CGCCATGGAGAGCATCGTGATCTGGGCGCAAGACCCCAAGTTCGGCTGCCCATCCAGAA 1844
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
577 eGlyLysGluAlaLeuValIleTrpGlyArgIleProLysPheHisLeuProValGluAr 597
QY 1845 GGAGACCTGGGACACCTGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGGA 1904
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
597 gGluIleTrpGluGlnTrpTrpAspAspTyrTrpGlnValThrTrpIleProAspTrpAs 617
QY 1905 GTTCGTGAACACCCCGCTGGTGAAGCTGTGTACCACTGGAGAAGGAGGCCATCAT 1964
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
617 pPheValSerThrProProLeuValArgLeuValPheAsnLeuValLysAspProIlePr 637
QY 1965 CGCGCGCGAGACCTTCTACGTGACCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGGC 2024
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
637 oGlyAlaGluThrPheTyrThrAspGlySerCysAsnArgGlnSerLysGluGlyLysAl 657
QY 2025 CGGCTACGTGACCGCGGCGCGGCGAGAGATCGTGAGCCTGACCGAGACCAACCA 2084
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
657 aGlyTyrValThrAspLysGlyArgAspLysValLysValLeuGluGlnThrThrAsnGl 677
QY 2085 GAAGACCGAGCTGCAGGCCATCCAGCTGGCCCTGCAGGACAGCGGAGGAGGTGAACAT 2144
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
677 nGlnAlaGluLeuGluAlaPheAlaMetAlaLeuThrAspSerGlyProLysAlaAsnIl 697
QY 2145 CGTGACCGAGCAGCCAGTACGCCCTGGCGATCATCCAGGCCCGCCGACAGAGCGAGAG 2204
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
697 eIleValAspSerGlnTyrValMetGlyIleValValGlyGlnProThrGluSerGluAs 717
QY 2205 CGAGCTGGTGAACCATCATCGACGAGCTGATCAAGAGGAGAGAGGTGTACTGTAGCTG 2264
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
717 nArgIleValAsnGlnIleIleGluMetIleLysLysGluAlaIleTyrValAlaTr 737
QY 2265 GGTGCGCGCCCAACAAGGGATCGCGGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGG 2324
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
737 pValProAlaHisLysGlyIleGlyGlnGlnGluValAspHisLeuValSerGlnGl 757
QY 2325 CATCCGCAAGGTGCTGTTCTGACCGGCATCGAT 2358
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
757 yIleArgGlnValLeuPheLeuGluLysIleGlu 768

```

Search completed: June 2, 2005, 05:09:55  
Job time : 107.734 secs

***This Page Blank (uspio)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 1, 2005, 20:20:55 ; Search time 208.158 Seconds  
(without alignments)  
12088.668 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 4555

Sequence: 1 gtcagccaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

MODEL=frame+ n2p.model -DEV=xlh  
-O=/cgn2\_1/USPO\_pool/US09610313/runat\_31052005\_155136\_15147/app\_query.fasta.1.7893  
-DB=UniProt\_03 -QWMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR=US09610313 @CGN 1.1 725 @runat\_31052005\_155136\_15147 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOJURY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: uniprot\_sprot.\*
- 2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4097	89.9	1427	2 Q9WF92	Q9wf92 human immun
2	4093	89.9	1427	2 Q9WF89	Q9wf89 human immun
3	4089	89.8	1428	2 Q9WF86	Q9wf86 human immun
4	3947	86.7	1427	2 Q9WF62	Q9wf62 human immun
5	3940	86.5	1429	2 Q6X4P8	Q6x4p8 human immun
6	3912	85.9	1425	2 Q6X4Q6	Q6x4q6 human immun
7	3905	85.7	1427	2 Q9WF96	Q9wf96 human immun
8	3901	85.6	1430	2 Q6X4R4	Q6x4r4 human immun
9	3892.5	85.5	1427	2 Q9WF71	Q9wf71 human immun
10	3888	85.4	1427	2 Q9WF60	Q9wf60 human immun
11	3885.5	85.3	1427	2 Q9WF65	Q9wf65 human immun
12	3884	85.3	1437	2 Q9WF77	Q9wf77 human immun
13	3876.5	85.1	1427	2 Q9WF74	Q9wf74 human immun
14	3870.5	85.0	1427	2 Q9ADK5	Q9adk5 human immun
15	3862	84.8	1425	2 Q9WF79	Q9wf79 human immun
16	3859	84.7	1427	2 Q9WF98	Q9wf98 human immun

17	3850.5	84.5	1429	2 Q9WF84	Q9wf84 human immun
18	3841	84.3	1428	2 Q6X4P0	Q6x4p0 human immun
19	3831.5	84.1	1427	2 Q9WF68	Q9wf68 human immun
20	3795.5	83.3	1432	2 Q8AD29	Q8ad29 human immun
21	3792	83.2	1433	2 Q8AD89	Q8ad89 human immun
22	3790	83.2	1434	2 Q8ADV1	Q8adv1 human immun
23	3788	83.2	1430	2 Q7SV36	Q7sv36 human immun
24	3784	83.1	1425	2 Q7SV20	Q7sv20 human immun
25	3778.5	83.0	1433	2 Q8ADN7	Q8adn7 human immun
26	3777.5	82.9	1433	2 Q8ADK5	Q8adk5 human immun
27	3775.5	82.9	1433	2 Q8AD87	Q8ad87 human immun
28	3772.5	82.8	1433	2 Q8ADG5	Q8adg5 human immun
29	3769.5	82.8	1435	2 Q9DQ33	Q9dq33 human immun
30	3768.5	82.7	1433	2 Q8ADE1	Q8ade1 human immun
31	3763.5	82.6	1431	2 Q8ADR1	Q8adr1 human immun
32	3762.5	82.6	1433	2 Q8ADCS	Q8adc5 human immun
33	3761.5	82.6	1433	2 Q8AE32	Q8ae32 human immun
34	3760.5	82.6	1433	2 Q8ADM1	Q8adm1 human immun
35	3760.5	82.6	1433	2 Q8ADU3	Q8adu3 human immun
36	3759.5	82.5	1437	2 Q41782	Q41782 human immun
37	3758	82.5	1434	2 Q8Q727	Q8q727 human immun
38	3758	82.5	1435	2 Q9WLJ3	Q9wlj3 human immun
39	3757.5	82.5	1433	2 Q8AD85	Q8ad85 human immun
40	3755.5	82.4	1433	2 Q7SP79	Q7spt9 human immun
41	3751.5	82.4	1433	2 Q8ADW7	Q8adw7 human immun
42	3751.5	82.4	1436	2 Q8AD93	Q8ad93 human immun
43	3749.5	82.3	1433	2 Q8AE23	Q8ae23 human immun
44	3748.5	82.3	1433	2 Q8ADR9	Q8adr9 human immun
45	3746.5	82.3	1433	2 Q8ADH3	Q8adh3 human immun

#### ALIGNMENTS

##### RESULT 1

Q9WF92	PRELIMINARY;	PRT;	1427 AA.
ID Q9WF92			
AC Q9WF92;			
DT 01-NOV-1999 (TrEMBLrel. 12, Created)			
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)			
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE Gag-pol polyprotein.			
DE Human immunodeficiency virus 1.			
OS Viruses; Retroid viruses; Retroviridae; Lentivirus.			
OX NCBI_TaxID=11676;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=92214383; PubMed=10196340;			
RA Novitsky V.A., Montano M.A., McLane M.P., Renjifo B., Vannberg F.,			
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,			
RA Essex M.;			
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency			
RT virus type 1 subtype C: a set of 23 full-length clones from			
RT Botswana."			
RL J. Virol. 73:4427-4432(1999).			
CC -1- SIMILARITY: Belongs to peptidase family A2.			
DR EMBL; AF110975; AAO17135.1; -			
DR HSSP; P04585; 1JLA.			
DR GO; GO:0019012; C:virion; IEA.			
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.			
DR GO; GO:0003677; F:DNA binding; IEA.			
DR GO; GO:0008907; F:integrase activity; IEA.			
DR GO; GO:0008233; F:peptidase activity; IEA.			
DR GO; GO:0004523; F:ribonuclease H activity; IEA.			
DR GO; GO:0003723; F:RNA binding; IEA.			
DR GO; GO:0003664; F:RNA-directed DNA polymerase activity; IEA.			
DR GO; GO:0005198; F:structural molecule activity; IEA.			
DR GO; GO:0016740; F:transferase activity; IEA.			
DR GO; GO:0008270; F:zinc ion binding; IEA.			
DR GO; GO:0015074; P:DNA integration; IEA.			
DR GO; GO:0006310; P:DNA recombination; IEA.			
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.			
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.			
DR GO; GO:0016032; P:viral life cycle; IEA.			

500	Db	LyseGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro	519
493	Qy	GGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAGGTGCGCAGTAC	552
520	Db	GlyysTrpLysProLysMetIleGlyIleGlyPheIleLysValArgGlnTyr	539
553	Qy	GACCAGATCTGATCGAGATCTCGGCAAGAAGCCATCGGCACCGTGTGATCGGCCCC	612
540	Db	AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro	559
613	Qy	ACCCCGTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTGGAATTC	672
560	Db	ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe	579
673	Qy	CCCATCAGCCCATCGAGACCGTCCGCTGAAGCTGAAGCCGCGCATGAGCGGCCCAAG	732
580	Db	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
733	Qy	GTCAAGCAGTGGCCCTCGACGAGGAGAAAGATCAAGGCCCTGACCGCCATCTGCGAGGAG	792
600	Db	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu	619
793	Qy	ATGAGAAGGAGGGAAGATCAACAAGATCGGCCCGAGAAACCCCTTCAACACCCCGCTG	852
620	Db	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal	639
853	Qy	TTGCGCATCAAGAGAGGACAGCACCAAGTGGCGCAAGCTGTGTGGACTTCCCGAGCTG	912
640	Db	PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
913	Qy	AACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCGCCACCCCGCGGCTG	972
660	Db	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679
973	Qy	AAGAAGAAGAGCGTGAACGCTGTGGAAGTGGGCGACGCTACTTTCAGGCTGCCCTG	1032
680	Db	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
1033	Qy	GAGGAGACTTCGCGAAGTACCGCTTCACCATCCCGAGCATCAACAAGAGACCCCC	1092
700	Db	AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro	719
1093	Qy	GGCATCCGCTACCAGTACAACGTGCTGCCCGAGGCTGGAAGGCGACCCCGAGCATCTTC	1152
720	Db	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe	739
1153	Qy	CAGAGCAGCATGACCAAGATCTGTGAGCCCTTCGCGCGCGCGCAACCCCGAGATCGTGATC	1212
740	Db	GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle	759
1213	Qy	TACCAG-----GCCCGCTGACGTGGGCGAGCGACTCGAGATCGGCCAGCACCGCGCC	1266
760	Db	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779
1267	Qy	AAGATCAGGAGCTGCGCAAGACACTGCTGCGCTGGGGCTTCACACCCCGCACAAGAAG	1326
780	Db	LysIleGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys	799
1327	Qy	CACCAGAAGGAGCCCGCTTCCTG-----CCCATCGAGCTGCACCCGACAAAGTGGACC	1380
800	Db	HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr	819
1381	Qy	GTGAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACACATCCAGAACTG	1440
820	Db	ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu	839
1441	Qy	GTGGCAGAGTGAACCTGGCGCAGCCACATCTACCCCGGCATCAAGGTGCGCAGCTGTC	1500
840	Db	ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys	859
1501	Qy	AAGCTGCTCGCGCGCCAAAGGCCCTGACCGACATCTGTGCGCTTCGACCGAGGAGCCGAG	1560





Db 360 LeuAlaGluAlaMetSerGlnAlaThrSerAlaAsnIleLeuMetGlnArgSerAsnPhe 379  
:::  
Qy 73 AAGGCCCCAGCGCATCATCAAGTGCCTCAACTCGCGCAAGAGGGCCATCGCCCGC 132  
Db 380 LysGlyProLysArgIleIleLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399  
Qy 133 AACTCCCGCCGCCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAGGGCCACAGATG 192  
Db 400 AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419  
Qy 193 AAGACTGCACCGAGCGCCAGCCAACTTCTTCGCGAGGACCTGCGCTTCCCCAGGGC 252  
Db 420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAspLeuAlaPheProGlnGly 439  
Qy 253 AAGGCCCGGAGTTCCCGAGCGAGCAGAACCGCGCCAAACAGCCCCACAGCGCGAGCTG 312  
Db 440 LysAlaArgGluPheProSerGluGlnAsnArgAlaAsnSerProThrSerArgGluLeu 459  
Qy 313 CAGGTGCGCGGCAACACCCCGCAGCGAGCGCGCGCGCGAGCGCCAGCCCTGAAC 372  
Db 460 GlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGluArgGlnGlyThrLeuAsn 479  
Qy 373 TTCCCCCGAGATCACCGTGGCAGCGCCCGCTGCTGAGCATCAAGGTGGCGCGCAGATC 432  
Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle 499  
Qy 433 AAGGAGCGCTGCTGGACACCGCGCCGACGACACCGTCTCGGAGGATGAGCCCTGCC 492  
Db 500 LysGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluGluMetSerLeuPro 519  
Qy 493 GSCAAGTGAAGCCCAAGATGATCGCGCATCGCGCGCTTCATCAAGTGGCGCAGTAC 552  
Db 520 GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr 539  
Qy 553 GACCAGATCCTGATCGAGATCTCGGCAAGAGGCGCATCGGCACCGTCTGATCGGCCCC 612  
Db 540 AspGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuIleGlyPro 559  
Qy 613 ACCCCGTGAACATCATGCGCGCGCAACATGTCGACCGAGCTGGGTGACCCCTGAACCTC 672  
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe 579  
Qy 673 CCCATCAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGCATGAGCGGCCCAAG 732  
Db 580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy 733 GTGAAGCAGTGGCCCTGTACCGAGGAGAGATCAAGGCGCTGACCGCCATCTGCGAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
Qy 793 ATGGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAGAACCCCTACACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
Qy 853 TTCGCCCATCAAGAAGAGACACCAAGTGGCGCAAGCTGGTGCCTTCGCGAGCTG 912  
Db 640 PheAlaIleLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAAAGCGCACCGAGACTTCTGGGAGGTGCAGCTGGGCATCCCCACCCCGCGCGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAAGAAGCGTGACCGTGTGACGTGGGAGCGCCTACTTCAGCGTGGCCCTG 1032  
Db 680 LysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGGACTCCGCAAGTACACCGCTTCCATCCCGCAGCATCAACAAACGAGACCCC 1092  
Db 700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
Qy 1093 GGCATCCCGTACAGTACAACTGCTGCGCCCGAGGCTGGAAGGCGAGCCCGCAGCATCTTC 1152  
:::

Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCCTGAGCCCTTCGCGCCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
Qy 1213 TACCAG-----GCCCGCCCTGTACGTGGCGAGCACTGTGGAGATCGGCAGACCGCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCAGGAGAGTGGCGCAAGCAGCATCTGCTGCGCTGGGGCTTCACACCCCGCAGCAAGAG 1326  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
Qy 1327 CACCAGAGGAGCGCCCGCTTCCCTG-----CCCATCGAGCTGCACCCCGACAAGTGGACC 1380  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1381 GTGCAGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAAAGCTG 1440  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1441 GTGGGCAAGCTGAATCGGCGCAGCATCTACCCCGGCATCAAGTGGCGCAGCTGTC 1500  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
Qy 1501 AAGCTGCTGCGCGCGCCCAAGCGCTGACGACATCGTCCCTGACCGAGGAGGCCGAG 1560  
Db 860 LysLeuLeuArgGlyThrLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
Qy 1561 CTGAGCTGGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACCGCGTGTACTACGAC 1620  
Db 880 LeuGluLeuAlaGluAsnArgGluLeuArgGluProValHisGlyValTyrTyrAsp 899  
Qy 1621 CCCAGAGAGACCTGCTGCGCGAGATCCAGAAAGCAGGGCCACGACCTGACCTACCGAG 1680  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTAGTACGCCAAGATGGCACCGCC 1740  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1741 CACACCAACGACGTGAAGCAGCTGACCGAGGCGCTGTCAGAGAAGATCGCCATGGAGAGCATC 1800  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959  
Qy 1801 GTGATCTGGGCGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
Qy 1861 TGGTGGACCGACTACTGGCAGGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCC 1920  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1921 CCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCGCGAGACCTTC 1980  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
Qy 1981 TACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCGCTGCTGACCCGAC 2040  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
Qy 2041 CGGGCGCGCGAGAGATCGTGGCTGACCGAGACCAACCAAGACCGAGCTGGAG 2100  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
Qy 2101 GCCATCCAGCTGGCCCTGTCAGGACAGCGCGCAGAGGTGAACATCGTGACCGAGCCGAC 2160  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2161 TACGCCCTGGGATCATCTCAGGCCCGAGCCCGCAAGAGCAGGAGCGAGCTGGTGAACGAG 2220  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099



Db 580 ProfileSerProIleGluThrValProValLysLeuLysProGlyWetAspGlyProLys 599  
QY 733 GTGAGCAGTGGCCCTGACCGAGGAGAGATCAAGGCCCTGACCGCATCTGCCAGGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu 619  
QY 793 ATGGAGAGAGGAGGGAAGATCACCAAGATCGGCCCGCGAGAACCCCTCAACAACCCCGCTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
QY 853 TTGCGCATCAAGAAGACAGACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTyrAspLysLeuValAspPheArgGluLeu 659  
QY 913 AACAGCCGACCCAGGACTTCTGGAGGTGACCTGGGCGATCCCGACCCCGCGGCTG 972  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
QY 973 AAGAAGAGAAGAGCGTGAACCTGCTGGACGTGGCGCGCCTACTTTCAGCGTGGCCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
QY 1033 GACGAGGACTTCCGCAAGTACACCGCTTACCATCCCGAGCATCAACAACGAGACCC 1092  
Db 700 AspGluAspPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
QY 1093 GCATCCGCTACAGTACACGCTGCTGCCGAGGCTGGAGGGCAGGCCCGAGCATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProSerIlePhe 739  
QY 1153 CAGAGCAGCATCAACAAGATCTGGAGCCCTTCCGCGCCGCAACCCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaArgAsnProGluIleValIle 759  
QY 1213 TACCAG-----GCCCCCTGTACGTGGCAGCGACCTGGAGATCGGCCAGCACCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
QY 1267 AGATCGAGGAGCTGGCAGACCTGCTGGCTGGGCTTACCACCCCGCAGCAAGAAG 1326  
Db 780 LysIleGluGluLeuArgLysHisLeuLeuArgTrpGlyPheThrValAsnAspIleGlnLys 799  
QY 1327 CACCAAGAGGAGCCCTTCCTG-----CCCATCGAGCTGCACCCCGCAAGTGGAGCC 1380  
Db 800 HisGlnLysGluProPheLeuThrMetGlyTyrGluLeuHisProAspLysTrpThr 819  
QY 1381 GTGACGCCCATCGAGTGGCCGAGAGGAGAGCTGGACCGTGAACGACATCCAGAGCTG 1440  
Db 820 ValGlnProIleGluLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
QY 1441 GTGGCAGCTGAATGGGCCGAGCCAGATCTACCCCGGCATCAAGTGGCGCAGCTGTGC 1500  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
QY 1501 AAGCTGCTCGCGGGCCCAAGCCCTGACGACATCGTCCCTGCACCGAGGAGCGCAG 1560  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY 1561 CTGGAGCTGGCCGAGAACCGCAGATCTCTCGCGAGCCCGTGCACCGCGTGTACTACGAC 1620  
Db 880 LeuGluLeuAlaGluAspArgGluIleLeuArgGluProValHisGlyValTyrTrpAsp 899  
QY 1621 CCCAGGAGACTGGTGGCCGAGATCCAGAGCAGGGCCACGACCATGGAGCTTACCGAG 1680  
Db 900 ProSerLysAspLeuValAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
QY 1681 ATCTACCGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGGCGCACCGCC 1740  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
QY 1741 CACACCAACGAGTGAAGCAGCTGACCGAGCGCGCTGCAGAGATCGCCATGAGAGCATC 1800  
Db 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaMetGluSerIle 959

QY 1801 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 960 VallietrpglyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
QY 1861 TGTGGACCGGACTACTGCGAGGCCACCTGGATCCCGAGTGGGAGTTCTGTTGAACACCC 1920  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTyrGluPheValAsnThrPro 999  
QY 1921 CCCCTGTGAAGCTGTGTACAGCTGGAGAAGGCCCATCATCGGCGCGCAGACCTTTC 1980  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleIleGlyAlaGluThrPhe 1019  
QY 1981 TACGTGGACGCGCGCGCAACCGGAGACCAAGATCGCAAGCGCGCTACGTGACCCGAC 2040  
Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
QY 2041 CGGGCCCGCGCAAGATCGTGAGCTGACCGAGACCCACCAACCAAGAGACCGAGCTGCAG 2100  
Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
QY 2101 GCCATCAGCTGGCGCTCGCAGACACGCGGACGAGGTGAACATCGTACCGCAGCCAG 2160  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
QY 2161 TACGCCCTGGCATCATCCAGCCCGCAGCCGACCAAGAGCAGAGCGAGCTGGTGAACCCAG 2220  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
QY 2221 ATCATCGAGCAGCTGATCAAGAAGAGAGTGTACTCTGAGCTGGTGGTCCCGCCACAG 2280  
Db 1100 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1119  
QY 2281 GGCATCGCGCGCAGCAGCAGATCGCAAGCTGTGAGCAAGGCGATCCGCMAGTGTCTG 2340  
Db 1120 GlyIleGlyLysAsnGluGlnIleAspLysLeuValSerLysGlyIleArgLysValLeu 1139  
QY 2341 TTCCTGGACGCGCATCGAT 2358  
Db 1140 PheLeuAspGlyIleAsp 1145  
RESULT 4  
Q9WF62  
ID Q9WF62 PRELIMINARY; PRT; 1427 AA.  
AC Q9WF62;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_taxid=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndong'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RA "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana."; RT  
RL J. Virol. 73:4427-4432(1999).  
CC -!- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110961; AAD17021.1; -.  
DR HSSP; Q70622; 1HVN.  
DR GO; GO:0019012; C:virus; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.



QY 1381 GTGAGCCATCGAGCTGCCGAGAGGAGAGCTGGACGTGAACGACATCCAGAGCTG 1440  
 DB 820 ValGlnProIleAsnLeuProGluLysGluSerTrpThrValAsnAspIleGlnLysLeu 839  
 QY 1441 GTGGCAAGCTGAACCTGGCCAGCCAGATCTACCCCGGCATCAAGGTGGCCAGCTGTGC 1500  
 DB 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuCys 859  
 QY 1501 AGCTGCTGGCGGCGCCAGGCCCTGACCGACATCGTCCCTGACCGAGAGGCCGAG 1560  
 DB 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
 QY 1561 CTGGAGCTGGCGAGAACCGCCAGATCTCTGCGGAGCCCGTCCACGGGTGTACTACGAC 1620  
 DB 880 LeuGluLeuAlaGluAsnArgGluLeuLysGluProValIleGlyValTyrTrpAsp 899  
 QY 1621 CCCAGCAAGGACCTGTGGCCGAGATCCAGAACGAGGCCGACGACGAGCTGACCTACCAG 1680  
 DB 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
 QY 1681 ATCTACGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATGGCGACCGCC 1740  
 DB 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
 QY 1741 CACACCACGAGCTGAAGCAGCTGACCGAGCCGTCGAGAGATCGCCATGGAGAGCATC 1800  
 DB 940 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleAlaThrGluSerIle 959  
 QY 1801 GTGATCTGGGCAAGACCCCAAGTTCCGCTGCCCATCCAGAGGAGACCTGGGAGACC 1860  
 DB 960 ValIleTrpGlyLysThrProLysPheArgLeuProlleGlnLysGluThrTrpGluThr 979  
 QY 1861 TGGTGACCGACTACTGGCAGCCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCC 1920  
 DB 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
 QY 1921 CCCCTGGTGAAGCTGTGTACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACCTTC 1980  
 DB 1000 ProLeuValLysLeuTrpTyrGlnLeuGluGluProIleValGlyAlaGluThrPhe 1019  
 QY 1981 TAGCTGGAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACCTGACCGAC 2040  
 DB 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1039  
 QY 2041 CGGGCGCGCAGAAGATCTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2100  
 DB 1040 ArgGlyArgGlnArgIleValSerLeuThrGluThrThrAsnGlnLysThrGluLeuGln 1059  
 QY 2101 GCATCCAGCTGGCCCTGCAGACACCGCGCAGGAGTCAACATCGTGACCGACGCGAC 2160  
 DB 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGlyValAsnIleValThrAspSerGln 1079  
 QY 2161 TAGCCCTGGGCATCATCCAGCCCGCCAGAGCGAGCGAGCGAGCTGGTGAACCCAG 2220  
 DB 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
 QY 2221 ATCATCGACGACTGATCAAGAAGGAGAGGTGTACTTGTAGTGGTGGTCCCGCCACAG 2280  
 DB 1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
 QY 2281 GGCATCGCGGCAAGCAGCAGATCGACAGCTGTGTGACGAGGCGATCCGCAAGGTGCTG 2340  
 DB 1120 GlyIleGlyGlyLeuGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
 QY 2341 TTCCTGGACGCGATCGAT 2358  
 DB 1140 PheLeuAspGlyIleAsp 1145  
 PRELIMINARY; PRT; 1429 AA.  
 Q6X4P8  
 AC Q6X4P8;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DE Gag-pol fusion polyprotein (Fragment).  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=14709249; DOI=10.1089/0892220377181220;  
 RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birk D.L.,  
 RA McCutchan F.E.;  
 RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
 in full genome analysis, is not sustained in subgenomic regions.";  
 RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Harris M.E., Birk D.L.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kim B., McCutchan F.E.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Maayan S., Zeira M.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AY255825; AAF76538.1; -;  
 DR HSP; P03367; 1A8G.  
 DR GO; GO:0019012; C:virion; IEA.  
 DR GO; GO:0004190; P:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:peptidase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0008270; F:zinc ion binding; IEA.  
 DR GO; GO:0015074; P:DNA integration; IEA.  
 DR GO; GO:0006310; P:DNA recombination; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.  
 DR GO; GO:0016032; P:viral life cycle; IEA.  
 DR InterPro; IPR000721; Gag\_p24.  
 DR InterPro; IPR001037; Integrase C.  
 DR InterPro; IPR003308; Integrase Zn N.  
 DR InterPro; IPR001995; Peptidase\_A2.  
 DR InterPro; IPR009007; Pept\_Aspartic.  
 DR InterPro; IPR001969; Pept\_Asp\_AS.  
 DR InterPro; IPR010999; Retrovir\_matrix.  
 DR InterPro; IPR000071; Retrovir\_p17.  
 DR InterPro; IPR008916; Retrov\_capsid\_C.  
 DR InterPro; IPR008919; Retrov\_capsid\_N.  
 DR InterPro; IPR002156; RNaseH.  
 DR InterPro; IPR001584; Rve.  
 DR InterPro; IPR000477; RVTse.  
 DR InterPro; IPR010659; RVT\_connect.  
 DR InterPro; IPR010661; RVT\_thumb.  
 DR InterPro; IPR001878; Znf\_CCHC.  
 DR Pfam; PF00540; Gag\_p17; 1.  
 DR Pfam; PF00607; Gag\_p24; 1.  
 DR Pfam; PF02022; Integrase\_Zn; 1.  
 DR Pfam; PF00075; RNaseH; 1.  
 DR Pfam; PF00665; rve; 1.  
 DR Pfam; PF00077; RVP; 1.  
 DR Pfam; PF00078; RVT\_1; 1.  
 DR Pfam; PF06815; RVT\_connect; 1.  
 DR Pfam; PF06817; RVT\_thumb; 1.  
 DR Pfam; PF00098; zF-CCHC; 2.  
 DR PRINTS; PR00939; C2HCZNFINGER.



```
|||||
978 ThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrpIleProAspTrpGluPhe 997
|||||
1909 GTGACACACCCCCCTGGTGAAGCTGGTACAGCTGGAGAGAGAGCCCATCATCGGC 1968
|||||
998 ValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGly 1017
|||||
1969 GCCGAGACCTTCTACGTGACCGCGCCGCCCAACCCGAGACCAAGATCGGCAAGCCGCG 2028
|||||
1018 ValGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGly 1037
|||||
2029 TACGTGACCGACCGCGCGCGCGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAG 2088
|||||
1038 TyrValThrAspArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLys 1057
|||||
2089 ACCGAGCTGCAGCCATCATCGTGGCCCTGCGAGGACAGCGGAGCGAGTGAACATCGTG 2148
|||||
1058 ThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleVal 1077
|||||
2149 ACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCGCCGACGAGCGAGCGAGCGAG 2208
|||||
1078 ThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGlu 1097
|||||
2209 CTGGTGAACACGATCATCGACGAGCTGATCAAGAAGAGAGAGCTGCTACTGCTGGGTG 2268
|||||
1098 LeuValAsnGlnIleIleGluGlnLeuIleArgLysGluArgValTyrLeuSerTrpVal 1117
|||||
2269 CCGGCCCAAGGGATCGCGCGCAACGAGCGATCGACAGCTGGTGAAGCAAGGCGATC 2328
|||||
1118 ProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIle 1137
|||||
2329 CCGAAGGTGCTCTTCTGACCGCATCGAT 2358
|||||
1138 ArgLysValLeuPheLeuAspGlyIleAsp 1147

RESULT 6
Q6X4Q6 PRELIMINARY; PRT; 1425 AA.
AC Q6X4Q6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Gag-pol fusion polypeptide (Fragment).
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14709249; DOI=10.1089/088922203771881220;
RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrari G., Birk D.L.,
RA McCutchan F.E.;
RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed
RT in full genome analysis, is not sustained in subgenomic regions.";
RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).
RN [2]
RP SEQUENCE FROM N.A.
RA Harris M.E., Birk D.L.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Kim B., McCutchan F.E.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Maayan S., Zeira M.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to peptidase family A2.
DR EMBL: AY255824; AAP76529.1; -.
DR HSP; Q72498; 1k6Y.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
```

```
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0015074; F:DNA integration; IEA.
DR GO; GO:0006310; F:DNA recombination; IEA.
DR GO; GO:0006278; F:proteolysis and peptidolysis; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_P24.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn_N.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR009007; Pept Aspartic.
DR InterPro; IPR001969; Pept Asp AS.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010659; RVT_connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZAFINGER.
DR PRINTS; PR00234; HIVMATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR PROSITE; PS00158; ZF_CCHC; 2.
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
FT NON_TER 1
SQ SEQUENCE 1425 AA; 161115 MW; 782315D4D5305E18 CRC64;

Alignment Scores:
Pred. No.: 8.2e-148 Length: 1425
Score: 3912.00 Matches: 730
Percent Similarity: 96.69% Conservative: 30
Best Local Similarity: 92.88% Mismatches: 22
Query Match: 85.88% Indels: 4
Db: 2 Gaps: 2

US-09-610-313b-32 (1-2457) x Q6X4Q6 (1-1425)
QY 13 ATGGCGGAGGCGCATGAGCGAGCGCCACGAGCGCCACATCTGATGCGAGCGCAACTTC 72
Db ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
358 LeuAlaGluAlaMetThrGlnAlaAsnAsnThrHisIleMetMetGlnArgAsnAsnPhe 377
QY 73 AAGGCGCCCAAGCGCATCATCAAGTGTTCCTCACTCGGCAAGAGGGCCACATCCCGCGC 132
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
378 LysGlyProLysArgIleValLysCysPheAsnCysGlyLysGluGlyHisLeuAlaArg 397
QY 133 AACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 192
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
398 AsnCysArgAlaProArgLysGlyCysGlyLysGlyCysGlyLysGlyHisGlnMet 417
QY 193 AAGGACTGCACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252
```







Db 640 PheAlaIleLysLysLysAspSerThrIlystrpàrgLysLeuValAspPheAraGluLeu 659  
 Qy 913 AACAGCGCACCCAGGACTTCTGGAGAGTGCAGCTGGGCATCCCCACCCCGCGGCTG 972  
 Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
 Qy 973 AAGAGAGAGAGAGCGTGCCTGCTGAGCTGGCGACGCTACTTTCAGGCTGCCCTG 1032  
 Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
 Qy 1033 GACGAGGACTTCCCGAAGTACACCGCTTCCACCATCCCGACATCAACAACGAGACCC 1092  
 Db 700 AspGluGlyPheAraGlyLysThrAlaPheThrIleProSerIleAsnAsnAlaThrPro 719  
 Qy 1093 GGATCCGCTACCACTACAACGCTGCTGCCAGGCTGGAGGCGAGCCCGACGATCTTC 1152  
 Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
 Qy 1153 CAGAGCATGATGACCAAGATCTCGAGCCCTTCCGCGCCGCGCAACCCCGAGATCGTGATC 1212  
 Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGlyIleValIle 759  
 Qy 1213 TACCAG-----GCCGCCCTTCTAGTGGGACGACCTGGAGATCGGCGACCGCGCC 1266  
 Db 760 TyrGlnTyrMetAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
 Qy 1267 AAGATCAGGAGCTGCGCAAGCACCTGCTGGCTGGGGCTTCCACACCCCGCAAGAAG 1326  
 Db 780 LysIleGluLeuAraGThrHisLeuLeuLysTrpGlyPheThrThrProAspLysLys 799  
 Qy 1327 CACCAAGAGAGCCCTTCTCTG-----CCCATCGAGTGCACCCCGCAAGTGGACC 1380  
 Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
 Qy 1381 GTGCAGCCCTCGAGCTGCCGAGAGAGAGCTGGACCTGAGACGATCCAGACCTG 1440  
 Db 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
 Qy 1441 GTGGCAAGCTGAACCTGGGCGAGCAGATCTACCCCGGCATCAAGTGGCGCAGCTGTC 1500  
 Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleLysValArgGlnLeuLys 859  
 Qy 1501 AAGTGTCTGGCGCGCCAGAGCCCTGACGACATCTGTCCTGACCGAGGAGCGGAG 1560  
 Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleIleProLeuThrGluGluAlaGlu 879  
 Qy 1561 CTGAGCTGCGGAGACCGGAGATCTCGCGAGCCGCTGACGCGCTGTACTAGAC 1620  
 Db 880 LeuGluLeuAlaGluAsnAraGluIleLysGluProValHisGlyValTyrTyrAsp 899  
 Qy 1621 CCCAGCAAGACCTGTGGCGGAGATCCAGAGCGGCGCACACCACTGACCTACCGAG 1680  
 Db 900 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyHisAspGlnTrpThrGln 919  
 Qy 1681 ATCTACGAGGCGCTTCAAGAACCTCAAGACCGGCAAGTACGCAAGATCGCACCGCC 1740  
 Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
 Qy 1741 CACACCAACGCTGAGCAGCTGACCGAGCGCTGAGAGATCGCCATCGGAGATC 1800  
 Db 940 HisThrAsnAspValLysGlnLeuAlaGluAlaValGlnLysIleThrMetGluSerIle 959  
 Qy 1801 GTGATCTGGGCGAAGACCCCAAGTCTCGCTGCCCATCAGAGGAGACCTGGAGACC 1860  
 Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluThr 979  
 Qy 1861 TGTGGACCGACTACTCGCAGGCGCACCTGGATCCCGGAGTGGAGTTCGTGAACACCC 1920  
 Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
 Qy 1921 CCCTGTGAGTGTGGTACCACTGAGAGAGAGCGCCATCATCTGGCGCGGAGACCTTC 1980  
 Db 1000 ProLeuValLysLeuTyrGlnLeuLysGluProIleValGlyAlaGluThrPhe 1019

Qy 1981 TACGTGACGCGCGCCGCAACCGCGAGACCAAGATCGCAAGCGCGGCTACGTGACCGAC 2040  
 Db 1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
 Qy 2041 CGGGCGCGGAGAGATCGTGAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAG 2100  
 Db 1040 ArgGlyArgGlnLysIleValSerLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1059  
 Qy 2101 GCCATCCAGCTGGCGCTCGAGGACGAGGAGGAGTGAACATCTGTCAGCAGCAGCCAG 2160  
 Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlyAlaGluValAsnIleValThrAspSerGln 1079  
 Qy 2161 TACGCTCGGCGCATCATCCAGGCGCCAGCAAGAGCGAGCGAGCTGGTGAACCCAG 2220  
 Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1099  
 Qy 2221 ATCATCCAGCAGCTGATCAAGAGGAGAGTGTACTGAGCTGGTGGCGCCCGCCACAG 2280  
 Db 1100 IleIleGluGlnLeuIleAsnLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
 Qy 2281 GGCAATCGGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGCATCCGCAAGTGTG 2340  
 Db 1120 GlyIleGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
 Qy 2341 TTCCTGCGCGCATCGAT 2358  
 Db 1140 PheLeuAspGlyIleAsp 1145  
 RESULT 8  
 Q6X4R4  
 ID Q6X4R4 PRELIMINARY; PRT; 1430 AA.  
 AC Q6X4R4; 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Gag-pol fusion polypeptide (Fragment).  
 OS Human immunodeficiency virus 1.  
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=14709249; DOI=10.1089/08922203771881220;  
 RA Harris M.E., Maayan S., Kim B., Zeira M., Ferrarri G., Birx D.L.,  
 RA McCutchan F.E.;  
 RT "A cluster of HIV type 1 subtype C sequences from Ethiopia, observed  
 RT in full genome analysis, is not sustained in subgenomic regions.";  
 RL AIDS Res. Hum. Retroviruses 19:1125-1133(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Harris M.E., Birx D.L.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Kim B., McCutchan F.E.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Maayan S., Zeira M.;  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to peptidase family A2.  
 DR EMBL; AY255823; AAP76521.1; --  
 DR HSP; P12497; I89D  
 DR GO; GO:0019012; C:virion; IEA.  
 DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0008907; F:integrase activity; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.





Pred. No.:	4.88e-147	Length:	1427
Score:	3892.50	Matches:	723
Percent Similarity:	96.44%	Conservative:	35
Best Local Similarity:	91.98%	Mismatches:	23
Query Match:	85.46%	Indels:	5
DB:	2	Gaps:	3
US-09-610-313B-32 (1-2457) x Q9WF71 (1-1427)			
QY	13	ATGGCGGAGGCGCATGAGCGCCAGCCAGCCCAACATCTCTGATCGAGGGGAGCAACTTC	72
DB	361	LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsnPhe	379
QY	73	AAGGGCCCCAAGCGATCATCAAGTGCTTCAACTGCGGCAAGAGGGCCACATCCCGCGC	132
DB	380	LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg	399
QY	133	AACTGCGCGCGCCCCCGCAAGAGGCTGCTGGAAGTGGCGCAAGAGGGCCACAGATG	192
DB	400	AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet	419
QY	193	AAAGGACTGCACCGAGCGCCAGCAACTTCTTCGCGAGGAGCTGGCCTTCCCGCAGGC	252
DB	420	LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly	439
QY	253	AAGGCGCGGAGCTTCCCCAGCGAGCAGAAACCGCGCCAAACAGCCCCCAGCGCGAGCTG	312
DB	440	GluAlaArgGluPheProSerGlnThrArgAlaAsnSerProThrSerArgGluLeu	459
QY	313	CAGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG	372
DB	460	GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyAsnPheAsn	479
QY	373	TTCCCCCAGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGTGGGGCGCGCATC	432
DB	480	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlyGlnIle	499
QY	433	AAGGAGGCGCTCTGTGACACCGCGCGCGAGCACACCGTGTGGAGGATGAGCTGCGCC	492
DB	500	ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro	519
QY	493	GCAGATGGAGCCCAAGATGATCGCGGCATCGCGCTTCATCAAGTGGCGCGCATAC	552
DB	520	GlyLysTrpLysProLysIleIleGlyIleGlyLysPheIleLysValArgGlnTrp	539
QY	553	GACCGATCTCTGAGATCTCGCGCAAGAGGCGCATCGGCACCGTCTGATCGCGCCC	612
DB	540	GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro	559
QY	613	ACCCCGCTGAACATCATCGCGCGCAACATGTCACCGAGCTGGGCTGCACCCCTGAACCTTC	672
DB	560	ThrProIleAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsnPhe	579
QY	673	CCCATCAGCCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGATGAGCGGCCCAAG	732
DB	580	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
QY	733	GTGAGCAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGCCATCTGCGAGGAG	792
DB	600	ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluGlu	619
QY	793	ATGGAGAGGAGGCAAGATCACCAAGATCGCGCCCGGAGAACCCCTACAACACCCCGCTG	852
DB	620	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProIleAsnThrProVal	639
QY	853	TTCCGCATCAAGAAGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCGCGGAGCTG	912
DB	640	PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
QY	913	AACAGCGCACCCAGGACTTCTGGAGGTGAGTGGGCGATCCCCCACCCCGCGGCGCTG	972
DB	660	AsnLysArgThrGlnAspPheTrpGluValGlnLeuLeuGlyIleProHisProAlaGlyLeu	679

QY	973	AAGAAGAAAGAGCGTACCGTGTGGACGCTGGCGCGACGCTACTTCAGCGTGCCTCTG	1032
DB	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTrpPheSerValProLeu	699
QY	1033	GACGAGGACTTCCGCAAGTACACCGCTTCCACCTCCCGAGCATCAACAACGAGACCCCC	1092
DB	700	AspGluSerPheArgLysTrpAlaPheThrIleProSerIleAsnAsnSerThrPro	719
QY	1093	GGCATCCGCTACCAAGTACAAGTGTGCCCCAGAGGCTGGAGGGCGAGCCCGAGCATCTTC	1152
DB	720	GlyIleArgTrpGlnTrpAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	739
QY	1153	CAGAGCAGCATGACCAAGATCCTCGAGCCCTTCGCGCGCCGCAACCCCGAGATCGTGATC	1212
DB	740	GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
QY	1213	TACGAG-----GCCCGCTGTAGTGGCGAGCACTGTGGAGATCGGCAGCACCGCGCC	1266
DB	760	TyrGlnTrpMetAspAspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgAla	779
QY	1267	AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGGCTTCACCCACCCCGACAAGAG	1326
DB	780	LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProAspLysLys	799
QY	1327	CACCAGAGGAGCGCCCTTCTCTG-----CCCATCGAGCTGCACCCCGACAAGTGGACC	1380
DB	800	HisGlnLysGluProProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThr	819
QY	1381	GTGACGCCCATCGAGCTGCCGAGAGGAGAGCTGGACCGTGAACGACATTCAGAGCTG	1440
DB	820	ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu	839
QY	1441	GTGGCGAGCTGAACCTGGCGCAGCGCATCTACCCCGGCATCAAGTGGCGCGCATGCGC	1500
DB	840	IleGlyLysLeuAsnTrpAlaSerGlnIleTrpProGlyIleArgValLysLeuCys	859
QY	1501	AAGTGTCTGCGCGCGCCCAAGCGCTGACCCGACATCTGTCGCCCTGACGAGGAGCGCGAG	1560
DB	860	LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu	879
QY	1561	CTGAGCTGGCGCGAAGACCGCGAGATCTGCGGAGCGCGCTGCGCGCGGTGTATCAGAC	1620
DB	880	LeuGluLeuAlaGluAsnArgGluLeuLysGluProValHisGlyValTrpTrpAsp	899
QY	1621	CCCGAGGAGACCTGTGGCGGAGATCCAGAGGAGGCGCGCACCGACCTGACCTACCAG	1680
DB	900	ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTrpGln	919
QY	1681	ATCTACCAAGGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCCAAGATGCGACCCGCC	1740
DB	920	IleTrpGlnGluProPheLysAsnLeuLysThrGlyLysTrpAlaLysMetArgThrAla	939
QY	1741	CACACCAAGCGTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCCATGAGAGCATC	1800
DB	940	HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleAlaMetGluCysIle	959
QY	1801	GTGATCTGGCGCAAGACCCCAAGTTCGCTGCGCCCATCCAGAGGAGACCTGGGAGACC	1860
DB	960	ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr	979
QY	1861	TGGTGGACCGACTACTGCGAGCGCCACCTGGATCCCGAGTGGGAGTTCTGTAACACCCGCC	1920
DB	980	TrpTrpThrAspTrpTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro	999
QY	1921	CCCTGTGTGAAGCTGTGTATCAGCTGGAGAGAGGCCCATCATCGCGCGCGAGACCTTC	1980
DB	1000	ProLeuValLysLeuTrpTrpGlnLysGluProIleAlaGlyAlaGluThrPhe	1019
QY	1981	TACGTGAGCGCGCGCCCAACCGCGAGACCAAGATCGCAAGGCGCGCTAGTGCACCCAG	2040
DB	1020	TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTrpValThrAsp	1039
QY	2041	CGGGCGCGGAGAGATCGTGAGCTGACCGAGACCCACCAACCAAGAGACCGAGCTGCAG	2100







-I- SIMILARITY: Belongs to peptidase family A2.

CC	EMBL; AF110962; AAD17030.1; -;				
DR	HSPF; Q70622; 1HVN.				
DR	GO: GO:0019012; C:varion; IEA.				
DR	GO: GO:0004190; F:aspartic-type endopeptidase activity; IEA.				
DR	GO: GO:0003677; F:DNA binding; IEA.				
DR	GO: GO:0008907; F:integrase activity; IEA.				
DR	GO: GO:0008233; F:peptidase activity; IEA.				
DR	GO: GO:0004523; F:ribonuclease H activity; IEA.				
DR	GO: GO:0003723; F:RNA binding; IEA.				
DR	GO: GO:0003964; F:RNA-directed DNA polymerase activity; IEA.				
DR	GO: GO:0005198; F:structural molecule activity; IEA.				
DR	GO: GO:0016740; F:transferase activity; IEA.				
DR	GO: GO:0008270; F:zinc ion binding; IEA.				
DR	GO: GO:0015074; P:DNA integration; IEA.				
DR	GO: GO:0006310; P:DNA recombination; IEA.				
DR	GO: GO:0006508; P:proteolysis and peptidolysis; IEA.				
DR	GO: GO:0006278; P:RNA-dependent DNA replication; IEA.				
DR	GO: GO:0016032; P:viral life cycle; IEA.				
DR	InterPro; IPR000721; Gag_p24.				
DR	InterPro; IPR001037; Integrase C.				
DR	InterPro; IPR003308; Integrase Zn N.				
DR	InterPro; IPR001995; Peptidase_A2.				
DR	InterPro; IPR009007; Pept_AspArtic.				
DR	InterPro; IPR001969; Pept_Asp_AS.				
DR	InterPro; IPR010999; Retrovir_matrix.				
DR	InterPro; IPR000071; Retrovir_p17.				
DR	InterPro; IPR008916; Retrov_capsid C.				
DR	InterPro; IPR008919; Retrov_capsid_N.				
DR	InterPro; IPR002156; RNaseH.				
DR	InterPro; IPR001584; Rve.				
DR	InterPro; IPR000477; RVTse.				
DR	InterPro; IPR010659; RVT connect.				
DR	InterPro; IPR010661; RVT_thumb.				
DR	InterPro; IPR001878; Znf_CCHC.				
DR	Pfam; PF00540; Gag_p17; 1.				
DR	Pfam; PF00607; Gag_p24; 1.				
DR	Pfam; PF02022; Integrase_Zn; 1.				
DR	Pfam; PF00075; RNaseH; 1.				
DR	Pfam; PF00665; rve; 1.				
DR	Pfam; PF00077; RVP; 1.				
DR	Pfam; PF00078; RVT_1; 1.				
DR	Pfam; PF06815; RVT_connect; 1.				
DR	Pfam; PF06817; RVT_thumb; 1.				
DR	Pfam; PF00098; zf_CCHC; 2.				
DR	SMART; SM00343; Znf_C2HC; 2.				
DR	PROSITE; PS00141; ASP_PROTEASE; 1.				
DR	PROSITE; PS01175; ASP_PROT_RETROV; 1.				
KW	AIDS; Aspartyl protease; Core protein; Hydrolase; Polyprotein;				
KW	Protease; RNA-directed DNA polymerase; Transferase.				
SQ	SEQUENCE 1427 AA; 161347 MW; 595F19865DE81D63 CRC64;				

## Alignment Scores:

Pred. No.:	9.26e-147	Length:	1427
Score:	3885.50	Matches:	724
Percent Similarity:	96.44%	Conservative:	34
Best Local Similarity:	92.11%	Mismatches:	23
Query Match:	85.30%	Indels:	5
DB:	2	Gaps:	3

US-09-610-313B-32 (1-2457) x Q9WF65 (1-1427)

QY	13	ATGCGCGAGCGCATGAGCCAGCCAGCCCAACATCTGTGATGCGCGCAACTTC	72
Db	361	LeuAlaGluAlaMetSerGln---ThrAsnThrAsnValMetMetGlnArgSerAsn	379
QY	73	AAGGGCCCCAAGCGCATCATCAAGTGTCTCAACTGCGCGCAAGAGGCGCCACATCGCCGC	132
Db	380	LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg	399
QY	133	AATGCGCGCGCCCCCGCAAGAGGGTGTGTGGAGTTCGGCAAGAGGGGCCACCATG	192

Db	400	AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet	419
QY	193	AAGCACTGCACCGAGCGCCAGCCCAACTTCTTCGCGAGAGCACTGCGCTTCCCCAGGCG	252
Db	420	LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly	439
QY	253	AAGCGCGCGAGTTCCCGCAGCAGAGCAACCGCGCCCAACAGCCCCACAGCGCGAGGTG	312
Db	440	GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGlnLeu	459
QY	313	CAGGTGCGCGCGCAACCCCCCGCAGGAGCGCCGCGCGCGCGCGCGCGCGCGCGCGCG	372
Db	460	GlnValArgGlyAspAsnProArgSerGluThrArgAlaGluGlyGlnGlyThrLeuAsn	479
QY	373	TTCCCCCAGATCACCTGTGCGAGCGCCCCCTGTGTGAGCATCAAGGTGGCGCGCAGATC	432
Db	480	PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle	499
QY	433	AAGAGCGCTGTGTGGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	492
Db	500	ArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValLeuGluAspIleAsnLeuPro	519
QY	493	GGCAAGTGGAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	552
Db	520	GlyLysTrpLysProLysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyr	539
QY	553	GACCAGATCTGATCGAGATCTGCGGCAAGAGCCCATCGGCGCGCTCTCATCGGCCCC	612
Db	540	GluGlnIleLeuIleGluIleCysGlyLysLysAlaIleGlyThrValLeuValGlyPro	559
QY	613	ACCCCGTGAACATCATCGCGCGCAACATGCTGCCAGCGTGGCTGCGTGCACCTGAACCTC	672
Db	560	ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnLeuGlyCysThrLeuAsn	579
QY	673	CCCATGAGCCCATCGAGACGTCGCGCGTGAAGTGAAGCCCGCGCGCGCGCGCGCGCGCG	732
Db	580	ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys	599
QY	733	GTGAAGCAGTGGCGCCCTGACCGAGAGAAAGATCAAGGCCCTGACCGCGCATCTCGAGGAG	792
Db	600	ValLysGlnTrpProLeuThrGluGluLysAlaLeuThrAlaIleCysGluGlu	619
QY	793	ATGGAGAAGGAGGCAAGATCACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	852
Db	620	MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTrpAsnThrProVal	639
QY	853	TTGCGCATCAAGAAGAAGGACAGCACCAAGTGGCGCGCAAGCTGTGGAGATTCGCGGAGCTG	912
Db	640	PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu	659
QY	913	AACAAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGCGCATCCCCACCGCGCGCGCTG	972
Db	660	AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu	679
QY	973	AAGAAGAAGAGCGGTGACCGCTGTGTGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCG	1032
Db	680	LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu	699
QY	1033	GACGAGGACTTCGCAAGTACACCGCGCTTCACCATCCCCAGCATCAACAGAGACCCCC	1092
Db	700	AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro	719
QY	1093	GGCATCCGCTACCAAGTCAACGTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1152
Db	720	GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe	739
QY	1153	CAGACGAGCATGACCAAGATCTCGAGCCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1212
Db	740	GlnSerSerMetIleLysIleLeuGluProPheArgThrLysAsnProAspIleValIle	759
QY	1213	TACAG-----GCCCGCTGTACGTGGGCGAGGACTGTGAGATCGCGCGCGCGCGCGCGCGCG	1266
Db	760	TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla	779

1267 AAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGCTTCCACCAACCCCGACCAAG 1326  
Db |||||  
780 LysIleGluGluLeuArgGluHisLeuLeuLysTrpGlyLeuThrThrProTyrLysLys 799  
QY |||||  
1327 CACCAGAAGAGCCCTTCCTG-----CCATCGAGCTGCACCCCGACCAAGTGGACC 1380  
Db |||||  
800 HisGlnLysGluProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
QY |||||  
1381 GTGCGACCCATCGAGTGGCCGAGAGAGAGCTGACCGTGAACGATCCAGAGCTG 1440  
Db |||||  
820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValHisAspIleGlnLysLeu 839  
QY |||||  
1441 GTGGCGAGCTCACTGGCGCCAGCCAGATCTACCCCGCATCAAGGTGGCCAGCTGTC 1500  
Db |||||  
840 ValGlyLysLeuAsnTrpAlaSerGlnIleTrpProGlyIleArgValLysHisLeuCys 859  
QY |||||  
1501 AAGCTGCTCGCGGGCCCAAGCCCTTGACCGATCTGTCGCCCTGACCGAGAGAGCCGAG 1560  
Db |||||  
860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluGluAlaGlu 879  
QY |||||  
1561 CTGAGCTGCGCGAGAACCGCGAGATCTGCGCGAGCCCGTGCACGGCGTGTACTACGAC 1620  
Db |||||  
880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrTrpAsp 899  
QY |||||  
1621 CCAGCAGGACCTGTGGCCGAGATCCAGAGAGCGGGCCAGCAGCTGGAGCTTACCAG 1680  
Db |||||  
900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
QY |||||  
1681 ATCTACAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTAGCCCAAGATGCGCACCGCC 1740  
Db |||||  
920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetargThrAla 939  
QY |||||  
1741 CACACCAAGCGTGAAGCAGCTGACCGAGCGCGTGCAGAAATCGCCATGAGAGCATC 1800  
Db |||||  
940 HisThrAsnAspValLysGlnLeuThrGluValValGlnLysIleThrMetGluSerIle 959  
QY |||||  
1801 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCCCTCCATCCAGAGAGACCTGGAGACC 1860  
Db |||||  
960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979  
QY |||||  
1861 TGGTGGACCGACTACTGGCAGCGCCACCTGGATCCCGAGTGGAGTTCGTGAACACCCCC 1920  
Db |||||  
980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
QY |||||  
1921 CCCCTGGTGAAGCTGTGTACAGCTGGAGAGAGAGCCCATCATCGCGCCGAGACCTTC 1980  
Db |||||  
1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
QY |||||  
1981 TACGTGGAGCGCGCCCAACCGCAGACCAAGATCGGCAAGCGCGCTACGTGACCGAC 2040  
Db |||||  
1020 TyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrAsp 1039  
QY |||||  
2041 CGGGCGCGCAGAGATCGTGACCTGACCGAGACCCAGACACCAACAGAGACCCAGCTGCAG 2100  
Db |||||  
1040 LysGlyArgGlnLysIleValProLeuThrGluThrThrAsnGlnArgAlaGluLeuGln 1059  
QY |||||  
2101 GCATCCAGCTGGCCCTCGAGACAGCGCGCAGCGAGTGAACATCGTGACCAAGCGCAG 2160  
Db |||||  
1060 AlaIleGlnLeuAlaLeuAsnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
QY |||||  
2161 TACGCGCTGGCATCATCAGCGCCAGCCCGACAGAGCGAGCGAGCGCTGGTGAACAG 2220  
Db |||||  
1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
QY |||||  
2221 ATCATCGACGAGCTGATCAAGAGAGAGAGTGTATCTGAGCTGGGTGCCCCCACAAG 2280  
Db |||||  
1100 IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
QY |||||  
2281 GGCATCGCGCGCAACGAGCAGATCCGACAGCTGGTGGAGCAAGGGCATCCGCAAGTGTG 2340  
Db |||||  
1120 GlyIleGlyGlyAsnGluValAspLysLeuValSerSerGlyIleArgLysValLeu 1139

QY 2341 TTCTGCGAGCGCATCGAT 2358  
Db |||||  
1140 PheLeuAspGlyIleAsp 1145  
RESULT 12  
Q9WF77  
ID Q9WF77 PRELIMINARY; PRT; 1437 AA.  
AC Q9WF77;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-NOV-2004 (TRENBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
CC 1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF110967; AAD17072.1; -.  
DR HSSP; P24736; INCP.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; F:DNA integration; IEA.  
DR GO; GO:0006310; F:DNA recombination; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR GO; GO:0006278; F:RNA-dependent DNA replication; IEA.  
DR GO; GO:0016032; P:viral life cycle; IEA.  
DR InterPro; IPR000721; Gag\_p24.  
DR InterPro; IPR001037; Integrase\_C.  
DR InterPro; IPR003308; Integrase\_Zn\_N.  
DR InterPro; IPR001995; Peptidase\_A2.  
DR InterPro; IPR009007; Pept\_Aspartic.  
DR InterPro; IPR001969; Pept\_Asp\_AS.  
DR InterPro; IPR010999; Retrovir\_matrix.  
DR InterPro; IPR000071; Retrovir\_p17.  
DR InterPro; IPR008916; Retrov\_capsid\_C.  
DR InterPro; IPR002156; RNaseH.  
DR InterPro; IPR001584; Rve.  
DR InterPro; IPR000477; RVise.  
DR InterPro; IPR010659; RVT\_connect.  
DR InterPro; IPR010661; RVT\_thumb.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF00540; Gag\_p17; 1.  
DR Pfam; PF00607; Gag\_p24; 1.  
DR Pfam; PF02022; Integrase\_Zn; 1.  
DR Pfam; PF00075; RNaseH; 1.  
DR Pfam; PF00665; rve; 1.  
DR Pfam; PF00077; RVP; 1.  
DR Pfam; PF00078; RVT; 1.  
DR Pfam; PF06815; RVT\_connect; 1.  
DR Pfam; PF06817; RVT\_thumb; 1.  
DR Pfam; PF00098; zf-CCHC; 2.  
DR PRINTS; PR00939; C2HCZNFINGER.  
DR PRINTS; PR00234; HIVMATRIX.  
DR SMART; SM00343; Znf\_C2HC; 2.

DR PROSITE; PS00141; ASP\_PROTEASE; 1.  
 DR PROSITE; PS00175; ASP\_PROT\_RETROV; 1.  
 DR PROSITE; PS00158; ZF\_CCHC; 2.  
 KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;  
 KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;  
 KW Zinc-finger.  
 SQ SEQUENCE 1437 AA; 161978 MW; 65DAA2562FC317A6 CRC64;

Alignment Scores:  
 Pred. No.: 1.06e-146 Length: 1437  
 Score: 3884.00 Matches: 732  
 Percent Similarity: 95.73% Conservative: 30  
 Best Local Similarity: 91.96% Mismatches: 20  
 Query Match: 85.27% Indels: 14  
 DB: 2 Gaps: 4

US-09-610-313B-32 (1-2457) x Q9WF77 (1-1437)

QY	13	ATGGCGGAGGCGATGAGCGAGCGCCACCGCCCAACATCTCTGATGCGAGCGAGCAACTTC	72
DB	360	LeuAlaGluAlaMetSerGlnAlaAsnSerValAlaAsnMetMetGlnLysSerAsnPhe	379
QY	73	AAGGSCCCAGCGCATCATCAAGTGTCTCAACTGCGCAAGGAGGCGCCATCGCCGCG	132
DB	380	LysGlyProArgArgAsnValLysCysPheAsnCysGlyLysGluGlyHisIleAlaLys	399
QY	133	AACTGCCGCGCCCGCCGCAAGAGGCTGCTGGAAGTCCGCAAGAGGCGCCACAGATG	192
DB	400	AsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet	419
QY	193	AAGACTGCACGAGCGCGCAGCCCAACTTCTTCGCGAGGACCTGCGCTTCCCGCAGGC	252
DB	420	LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly	439
QY	253	AAGCGCGCGAGTCCCGAGCGAGCAGCAACCGCGCC-----AACAGC	294
DB	440	GluAlaArgGluPheProGlnGlnIleArgAlaSerSerProAsnSerThrAsnSer	459
QY	295	CCACACGCGAGCTCAGTGGCGCGGCGACCAACCCCGCAGCGAGCGCGCGCGCGAG	354
DB	460	ProThrSerArgGluLeuGlnValArgGlyAspAsnProArgSerGluAlaGlyAlaGlu	479
QY	355	-----CGCCAGGCGACCTGAACTTCCCGCAGATCACCTGTGCGAGCGGCC	402
DB	480	GlyGlnGlyThrLeuGlnGlyThrLeuAsnCysProGlnIleThrLeuTrpGlnArgPro	499
QY	403	CTGTGACATCAAGTGGCGCGCGAGATCAAGAGGCGCTGTGGACACCGCGCGCGAC	462
DB	500	LeuValSerIleLysValGlyGlnIleLysGluAlaLeuLeuAspThrGlyAlaAsp	519
QY	463	GACACCGTGTGGAGGATGAGCTGCGCGCGAGTGGAAGCCCAAGATCATCGCGGC	522
DB	520	AsnThrValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGly	539
QY	523	ATCGCGCGCTTCATCAAGGTGCGCGAGTACACACAGATCTCTGATCGATCTGGGCAAG	582
DB	540	IleGlyGlyPheIleLysValArgGlnTrpAspGlnIleValIleGluIleCysGlyLys	559
QY	583	AAGCCATCGGACCGCTGATCGCGCGCCACCCCGTGAACATCATCGCGCGCAACATG	642
DB	560	LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet	579
QY	643	CTGACCGAGTGGGTGACCTGAACTTCCCATCAGCCCGCCNCGAGACCGTCCCGTG	702
DB	580	LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal	599
QY	703	AAGTGAAGCCCGCATCGAGCGCGCCCAAGGTGAAGTGAAGTGGCCCTCAGCCGAGAGAAG	762
DB	600	LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpThrLeuThrGluGluLys	619
QY	763	ATCAGGCGCTGACCGCATCTCGGAGGAGATGGAGAAGGAGGCGCAAGATCACCAAGATC	822
DB	620	IleLysAlaLeuThrGluIleCysGluGluMetGluLysGluGlyLysIleThrLysIle	639

QY	823	GGCCCCGAGAACCCCTTACAACACCCCGTGTTCGCCATCAAGAAGAGGACAGACCAAG	882
DB	640	GlyProGluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLys	659
QY	883	TGGCGCAAGCTGTGTGGACTTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTG	942
DB	660	TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal	679
QY	943	CAGCTGGCGATCCCGCCACCCCGCGCTGAAGAAGAGAGAGCGTGCCTGTCGAC	1002
DB	680	GlnLeuGlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAsp	699
QY	1003	GTGGCGACGCTTACTTTCAGCGTCCCGCTGAGCAGGACTTCCCGAGTACACCGCTTC	1062
DB	700	MetGlyAspAlaTyrPheSerValProLeuAspGluGlyPheArgLysThrAlaPhe	719
QY	1063	ACCATCCCGAGCATCAACAACGAGACCCCGCATCCGCTACCATAGTCAACGCTGCTGCC	1122
DB	720	ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTrpAsnValLeuPro	739
QY	1123	CAGGCTGGAAGGCGAGCCCGCAGCATCTTCAGAGCAGCATGACCAAGATCCTGGAGGCC	1182
DB	740	GlnGlyTrpLysGlySerProAlaIlePheGlnSerMetThrLysLysLeuGluPro	759
QY	1183	TTCCGCGCCCGCAACCCCGAGATCGTATCTACAG-----GCCCGCTGTACGTGGGC	1236
DB	760	PheArgLeuGlnAsnProGluIleValIleTyrGlnTrpMetAspAspLeuTyrValGly	779
QY	1237	AGCGACCTGGAGATCGCGCGCAGCACCGCCAGATCGAGGAGCTGCGCAGCAGCTGCTG	1296
DB	780	SerAspLeuGluIleGlyGlnArgAlaGlnIleGluGluLeuArgGluHisLeuLeu	799
QY	1297	CGCTGGGCGCTTCCACCCCGCAAGAAGCAGCAGAGGAGCGCCCTCTCTG-----	1350
DB	800	LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPheLeuTrpMet	819
QY	1351	CCCATCGAGCTGCAACCCCGCAAGTGGACCGTGCAGCCCATCGAGTGCCTCGAAGAGG	1410
DB	820	GlyTyrGluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysAsp	839
QY	1411	AGCTGGACGTGAACGACATCCAGAACCTGTGGCGAGCTGAACCTGGCGCAGCAGATC	1470
DB	840	SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle	859
QY	1471	TACCCCGCATCAAGGTGCGCGACGTGTGCAAGCTGTGCGCGCGCCCAAGCGCTGACC	1530
DB	860	TyrProGlyIleLysValArgGlnLeuLysLeuLeuArgGlyAlaLysAlaLeuThr	879
QY	1531	GACATCGTGGCGCGCGAGCGCGAGCTGCGAGCTGCGCGAGAACCGCGAGATCCTG	1590
DB	880	AspValValProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeu	899
QY	1591	CGCGAGCCCGTGCAGCGCGTGTACTACGACCCCGAGAGAGCTGTGGTGGCGAGATCCAG	1650
DB	900	LysGluProValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGln	919
QY	1651	AAGCAGGCGCCACGACCTGACCTTACAGATCTACAGGAGCGCTTCAAGAACCTGAAG	1710
DB	920	LysGlnGlyHisAspGlnTrpThrTyrGlnIleTyrGlnGluProPheLeuAsnLeuLys	939
QY	1711	ACCGGCAAGTACGCAAGATGCGCACCGCCCGCACCAACGACGAGTGAAGCAGCTGACCGAG	1770
DB	940	ThrGlyLysTyrAlaLysMetArgThrAlaHisThrAsnAspValLysGlnLeuThrGlu	959
QY	1771	GCCTGTCAGAAAGTCCCATGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCGCG	1830
DB	960	AlaValGlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArg	979
QY	1831	CTGCGCATCCAGAGGAGACCTGGGAGACCTGGTGGAGCCGACTACTGCGAGCCACCTGG	1890
DB	980	LeuProIleGlnLysGluThrTrpGluThrTrpTrpThrAspTyrTrpGlnAlaThrTrp	999

```
QY 1891 ATCCCGAGTGGAGTTGTTGAACACCCCGCCCTGGTGAAGCTGTGGTACCAGCTGGAG 1950
Db 1000 lIeProGluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGlu 1019
QY 1951 AAGGAGCCCATCATCGCCCGCCAGACCTTCTACTGTGGACCGCGCCGCAACCGCGAGACC 2010
Db 1020 LysGluProIleProGlyValGluThrPheTyrValAspGlyAlaAsnArgGluThr 1039
QY 2011 AAGATCGGCAAGCGCGGTACTGTGACCGACCGCGCGCGCGCAGAAGATCGTGAAGCTGACC 2070
Db 1040 LysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysIleValSerLeuThr 1059
QY 2071 GAGACCAACACAGAGACCGAGCTGTCAGGCGATCCAGCTGGCCCTGCAGGACGCGC 2130
Db 1060 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGly 1079
QY 2131 AGCGAGGTGAACATCGTACCCAGACGACAGTACGCGCTGGGCATCATCCAGGCCCGAGCCC 2190
Db 1080 SerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnPro 1099
QY 2191 GACAGAGCGAGGAGCGAGTGTGTGAACCAAGATCATCGACGAGCTGATCAAGAGGAGAAG 2250
Db 1100 AspLysSerGluSerGluLeuValAsnGlnIleIleGluGlnLeuIleGlnLysGluTrp 1119
QY 2251 GTGTACCTGAGCTGGTGGCCCGCCACAGGCGATCGCGCGCAACAGCAGCATCGACAG 2310
Db 1120 ValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLys 1139
QY 2311 CTGGTGAGCAAGGCGATCGCAAGTGTCTGTCTCGGACGCGCATCGAT 2358
Db 1140 LeuValSerGlnGlyIleArgLysMetLeuPheLeuAspGlyIleAsp 1155

RESULT 13
Q9WF74 ID Q9WF74 PRELIMINARY; PRT; 1427 AA.
AC AT
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Gag-pol polyprotein.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99214383; PubMed=10196340;
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,
RA Essex M.,
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency
RT virus type 1 subtype C: a set of 23 full-length clones from
RT Botswana.";
RL J. Virol. 73:4427-4432(1999).
CC -!- SIMILARITY: belongs to peptidase family A2.
DR EMBL; AF110965; RAD17056.1; -.
DR PIR; S49086; S49086.
DR HSP; Q70622; 1HVN.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008907; F:integrase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004523; F:ribonuclease H activity; IEA.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0015074; P:DNA integration; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006278; P:RNA-dependent DNA replication; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
```

```
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR001037; Integrase_C.
DR InterPro; IPR003308; Integrase_Zn_N.
DR InterPro; IPR001995; Peptidase_A2.
DR InterPro; IPR009007; Pept_AspArtic.
DR InterPro; IPR001969; Pept_Asp_AS.
DR InterPro; IPR010999; Retrovir_matrix.
DR InterPro; IPR000071; Retrovir_p17.
DR InterPro; IPR008916; Retrov_capsid_C.
DR InterPro; IPR008919; Retrov_capsid_N.
DR InterPro; IPR002156; RNaseH.
DR InterPro; IPR001584; Rve.
DR InterPro; IPR000477; RVTse.
DR InterPro; IPR010859; RVT_connect.
DR InterPro; IPR010661; RVT_thumb.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF02022; Integrase_Zn; 1.
DR Pfam; PF00075; RNaseH; 1.
DR Pfam; PF00665; rve; 1.
DR Pfam; PF00077; RVP; 1.
DR Pfam; PF00078; RVT_1; 1.
DR Pfam; PF06815; RVT_connect; 1.
DR Pfam; PF06817; RVT_thumb; 1.
DR Pfam; PF00098; zF_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIV1MATRIX.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS00175; ASP_PROT_RETROV; 1.
DR PROSITE; PS50158; zF_CCHC; 2.
KW AIDS; Aspartyl protease; Core protein; Hydrolase; Metal-binding;
KW Polyprotein; Protease; RNA-directed DNA polymerase; Transferase; Zinc;
KW Zinc-finger.
SQ SEQUENCE 1427 AA; 161299 MW; 924682CF5095D763 CRC64;

Alignment Scores:
Pred. No.: 2,11e-146 Length: 1427
Score: 3876.50 Matches: 720
Percent Similarity: 96.31% Conservative: 37
Best Local Similarity: 91.60% Mismatches: 24
Query Match: 85.10% Indels: 5
DB: 2 Gaps: 3

US-09-610-313B-32 (1-2457) x Q9WF74 (1-1427)
QY 13 ATGGCGGAGGCGCATGAGCCAGCGCCAGCCACATCTGTGTCAGCGCAGCACTTC 72
Db 361 LeuAlaGluAlaMetSerGlnAla---AsnThrSerValMetMetGlnLysSerAsnPhe 379
QY 73 AAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGGAGGGCCACATCCCGCGC 132
Db 380 LysGlyProArgArgIleValLysCysPheAsnCysGlyLysGluGlyHisIleAlaArg 399
QY 133 AACTGCGCGCGCCCGCCCAAGAGCGCTGTGGAGTGTGGAGTGGCGGCAAGGCCACCATG 192
Db 400 AsnCysArgAlaProArgLysGlyCysTrpLysCysGlyLysGluGlyHisGlnMet 419
QY 193 AAGGACTGCACCGAGCGCCAGCCCAACTTCTTCGCGAGGAGCCTGGCCCTCCCCAGGCGC 252
Db 420 LysAspCysThrGluArgGlnAlaAsnPhePheArgGluAsnLeuAlaPheProGlnGly 439
QY 253 AAGCGCGCGAGTGTCCCGAGGAGCAGAACCGCGCCAGCCAGCCAGCCAGCGCGAGCTG 312
Db 440 GluAlaArgGluPheProSerGluGlnThrArgAlaAsnSerProThrSerArgGluLeu 459
QY 313 CAGGTGCGCGCGCAACACCCCGCAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 372
Db 460 GlnValArgGlyAspAsnProArgSerGluThrArgValGluGlyGlnGlyAsnPheAsn 479
QY 373 TTCCCCCAGATCACCTGTGTGGCAGCGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 432
```

Db 480 PheProGlnIleThrLeuTrpGlnArgProLeuValSerIleLysValGlyGlnIle 499  
Qy 433 AAGGAGGCGCTGCTGGACACCGCGCGCGACACACCGCTGCTGGAGAGATGAGCTGCC 492  
Db 500 ArgGluAlaLeuLeuAaspThrGlyAlaAaspThrValLeuGluAaspIleAsnLeuPro 519  
Qy 493 GGCAGTGGAAAGCAAGATGATCGGCGGCATCGCGGCTTTCATCAAGTGGCGCAGTAC 552  
Db 520 GlyLysTrpLysProLysMetIleGlyIleGlyPheIleLysValArgGlnTyr 539  
Qy 553 GACCAGATCCTGATCGAGATCTGGCGCAAGAGCCATCGCACCGTCTGATCGGCGCC 612  
Db 540 GluGlnIleuIleGluIleCysGlyLysAlaIleGlyThrValLeuValGlyPro 559  
Qy 613 ACCCGCTGAACATCATCGCGCGCAACATGCTGACCCAGCTGGGCTGCACCTCAACTTC 672  
Db 560 ThrProValAsnIleIleGlyArgAsnMetLeuThrGlnPheGlyCysThrLeuAsnPhe 579  
Qy 673 CCATACAGCCCATCGAGACCGTCCGCTGAAGCTGAAGCCCGCATGAGCGGCCCAAG 732  
Db 580 ProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGlyProLys 599  
Qy 733 GTGAAGCAGTGGCCCTGACCGAGAGAGATCAAGCCCTGACCGCATCTCGAGAG 792  
Db 600 ValLysGlnTrpProLeuThrGluGluLysIleLysAlaLeuThrAlaIleCysGluAasp 619  
Qy 793 ATGAGAGAGGAGGCAAGATCACCAAGATCGCGCCCGAGAACCTTACACACACCCCGTG 852  
Db 620 MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProVal 639  
Qy 853 TTCGCCATCAAGAAAGAGCAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTG 912  
Db 640 PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
Qy 913 AACAGGCGCCACAGGACTTCTGGAGTGTGGAGTGGCGCTACTTCAAGCTGCCCTG 1032  
Db 660 AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
Qy 973 AAGAAGAAAGAGCGTGCCTGCTGACCTGGCGGACGCTACTTCAAGCTGCCCTG 1032  
Db 680 LysLysLysLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
Qy 1033 GACGAGGACTTCCGAAGTACACCGCTTCCACCATCCCGAGCATCAACACGAGACCC 1092  
Db 700 AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnSerThrPro 719  
Qy 1093 GGCATCGCTACAGTACAGCTGCTGCCCGAGGCTGGAGGCGAGCCCGAGATCTTC 1152  
Db 720 GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
Qy 1153 CAGAGCAGCATGACCAAGATCTCGAGCCCTTCCCGCGCCGACACCCGAGATCGTGATC 1212  
Db 740 GlnSerSerMetThrLysIleLeuGluProPheArgThrLysAsnProAspIleValIle 759  
Qy 1213 TACCAG-----GCCCGCTGTACGTGGCGAGCAGCTCGAGATCGGCGCACCGCGCC 1266  
Db 760 TyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAla 779  
Qy 1267 AAGATCAGAGGCTGCGGAGCAGCTGCTGGCTGGGCTTTCACACCCCGCAGCAAGAG 1326  
Db 780 LysIleGluLeuLeuArgGluHisLeuLeuLysTrpGlyLeuThrTrpProAspLysLys 799  
Qy 1327 CACCAGAGGAGCGCCCTTCCTG-----CCCATGAGCTGCACCCCGACAGTGAGCC 1380  
Db 800 HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
Qy 1381 GTGAGCGCATCGAGCTGCCGAGAGAGAGCTGGACCGCTGAACGATCCAGCATCCAGAGCTG 1440  
Db 820 ValGlnProIleGlnLeuProAspLysAspSerTrpThrValAsnAspIleGlnLysLeu 839  
Qy 1441 GTGGCGAAGCTGAATCTGGCGCAGCAGATCTACCCCGGCTCAAGAGTGGCGCAGCTGTC 1500  
Db 840 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValArgHisLeuCys 859

Qy 1501 AAGTGTGTGGCGCGCCAAAGCGCCCTGACCGACATCGTGGCCCTGACCGAGGAGCGCAG 1560  
Db 860 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu 879  
Qy 1561 CTGGAGTGGCGGAGAACCGCGAGATCTCTGGCGAGCCCGTGCACGGGTGTACTACGAC 1620  
Db 880 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 899  
Qy 1621 CCAGCAAGACCTGTGGCGCGAGATCCAGAAGCAGGCGCCACGACCTAGTGAGCTACCCAG 1680  
Db 900 ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyHisAspGlnTrpThrTyrGln 919  
Qy 1681 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGCGCAAGTACGCGCAAGATGCGCACGCC 1740  
Db 920 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 939  
Qy 1741 CACACCAACGACGTGAAGCAGTGCACCGAGCCCGTGCAGAGATCGGCATCGGAGCATC 1800  
Db 940 HisThrAsnAspValLysGlnLeuThrGluValValArgLysIleAlaMetGluSerIle 959  
Qy 1801 GTGATCTGGGCGAGACCCCAAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACC 1860  
Db 960 ValIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysAspThrTrpGluThr 979  
Qy 1861 TGTGAGACCGACTACTGCGCAGGCGCACCTGGATCCCGAGTGGGAGTTCGTGAACACACCCC 1920  
Db 980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
Qy 1921 CCCTGTGTGAAGCTGTGTACCTGAGTGGAGAGAGCCCATCATCTGGCGCGCAGACCTTC 1980  
Db 1000 ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleAlaGlyAlaGluThrPhe 1019  
Qy 1981 TAGCTGACCGCGCGCCACCGCAGACCAAGATCGGCGAGCGCGCTAGCTACCGAC 2040  
Db 1020 TyrValAspGlyAlaAlaAsnArgLysLeuGlyLysAlaGlyTyrValThrAsn 1039  
Qy 2041 CGGCGCGCGCAGAGATCGTGAGCTGAGCTGAGCAGACCAACCAAGAGACCGAGCTGCAG 2100  
Db 1040 LysGlyArgGlnLysIleValProLeuThrGluThrAsnGlnLysAlaGluLeuGln 1059  
Qy 2101 GCATTCAGTGTGGCCCTGCGAGCAGCGCAGGAGTGAACTCATCTGAGCCGACGAC 2160  
Db 1060 AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGln 1079  
Qy 2161 TAGCCCTGGGCATCATCCAGGCGCCCGCAGACCAAGAGCGAGCGAGCTGGTGAAC 2220  
Db 1080 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValThrGln 1099  
Qy 2221 ATCATCGAGCAGCTGATCAAGAAGAGAGAGTGTACCTGAGCTGGGTGGCGCCCAAG 2280  
Db 1100 IleIleGluGlnLeuIleLysLysGluArgIleTyrLeuSerTrpValProAlaHisLys 1119  
Qy 2281 GGCATCGCGCGCAACGAGCAGATCGACAAGCTGGTGAAGGCGCATCCGCAAGGTGCTG 2340  
Db 1120 GlyIleGlyAsnGluValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
Qy 2341 TTCTGCGACCGCATCGAT 2358  
Db 1140 PheLeuAspGlyIleAsp 1145

RESULT 14  
Q8ADX5 PRELIMINARY; PRT; 1427 AA.  
AC Q8ADX5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol fusion polypeptide (Fragment).  
GN Name=gag-pol;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;





793 ATGAGAGGAGGCGAAGATCACCAAGATCGGCCCCCGAGAACCCCTACACACACCCCGGTG 852  
Db MetGluLysGluGlyLysIleThrLysIleGlyProGluAsnProTyrAsnThrProIle 639  
853 TTCGCCATCAAGAAAGAGCAGCAGCAAGTGGCGCAAGCTGGTGGATTCCTCCGAGCTG 912  
Db PheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeu 659  
913 AACAAAGCGCACCCAGGACTTCTGGAGGTGAGCTGGCGCATCCCGCCCGCGGCTG 972  
Db AsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeu 679  
973 AAGAAGAAAGACGTGACCGTCTGTGAGTGGCGACCGCTACTTACGCGTCCCGCTG 1032  
Db LysLysArgLysSerValThrValLeuAspValGlyAspAlaTyrPheSerValProLeu 699  
1033 GACGAGACTTCCCAAGTACACCGCTTACCATCCCGAGCATCAACAGCAGACCCCG 1092  
Db AspGluSerPheArgLysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrPro 719  
1093 GGATCCGCTACCAAGTACAGCTGCTGCCAGGGCTGGAGGCGAGCCCGCAGCATCTTC 1152  
Db GlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePhe 739  
1153 CAGAGCAGCATGACCAAGATCCTGAGCCCTTCCGCGCCCGCAACCCCGAGATCGTGATC 1212  
Db GlnAsnSerMetThrLysIleLeuGluProPheArgAlaGlnAsnProGluIleValIle 759  
1213 TACCAG-----GCCCGCTGTAGTGGCGAGCAGCTGGAGATCGCGCAGCCGCGCC 1266  
Db TyrGlnTyrMetAspLeuTyrValGlySerAspLeuAspIleGluGlnHisArgAla 779  
1267 AAGATCGAGGAGCTGCGCAAGCAGCTGCTGGCTGGGGCTTCCACCCCGCCCGCAAGAA 1326  
Db LysIleGluLeuArgGluHisLeuLeuArgTrpGlyPheThrThrProAspLysLys 799  
1327 CACCAAGAGGAGCCCGCTTCTCTG-----CCCATCAGCTGCACCCCGCAAGTGAAC 1380  
Db HisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThr 819  
1381 GTGAGCCCATCGAGCTGCGGAGAGGAGGAGCTGGACCTGACGACATCCAGCATCCGAGCTG 1440  
Db ValGlnProIleGlnLeuProGluLysAspSerTrpValAsnAspIleGlnLysLeu 839  
1441 GTGGCAAGCTGAACCTGGCGCAGCAGCATCTACCCCGCATCAAGTGGCGCAGCTGTC 1500  
Db ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyLysValLysGlnLeuLys 859  
1501 AAGCTGTCGCGCGCGCAAGGCGCTGACCGACATCGTGGCCCTGACCGAGGAGCGCGAG 1560  
Db LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProLeuThrGluAlaGlu 879  
1561 CTGAGCTGGCGAGACCGGAGATCTCGCGGAGCCCTGCGAGCGCTGACCGGCTGACTAGAC 1620  
Db LeuGluLeuAlaGluAsnAsgluLeuLysGluProValHisGlyValTyrTyrAsp 899  
1621 CCACAGCAAGCACTGTGTGGCGGAGATCCAGAAAGCAGGCGCCACCGACCATGACCTACCAG 1680  
Db ProSerLysAspLeuIleAlaGluValGlnLysGlnGlyGlnAspGlnTrpTyrGln 919  
1681 ATCTACGAGGAGCCCTTCAAGAACTGAAGACCGCGCAAGTACGCGCAAGATGCGCAGCC 1740  
Db IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysValArgThrAla 939  
1741 CACACCAACGCTGAGCAGCTGACCGAGCGCTGCAAGATCGCGCATGGAGGATC 1800  
Db HisThrAsnAspIleLysGlnLeuThrGluAlaValGlnLysIleAlaArgGluSerIle 959  
1801 TGATCTGGGCGCAGACCCCAAGTCTCGCTGCGCTCATCAGAAAGGAGAGACTGGAGACC 1860  
Db IleIleTrpGlyLysThrProLysPheArgLeuProIleGlnLysGluThrTrpGluAla 979  
1861 TGGTGGACCGACTACTGGCAGGCGCACCTGGATCCCGAGTGGGAGTTCTGTGAACACCCCG 1920

980 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 999  
1921 CCCCTGTGTAAGCTGTGCTACCGAGTGAAGAGAGCCCATCATCGGCGCGAGACCTTC 1980  
Db ProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleProGlyValGluThrPhe 1019  
1981 TAGCTGACGCGCGCCCAACCGCAGACCAAGATCGCAAGCGCGCTACGTGACCGAC 2040  
Db TyrValAspGlyAlaAlaAsnArgGluThrLysMetGlyLysAlaGlyTyrValThrAsp 1039  
2041 CGGGCGCGCAGAGATCTGTAGCCTGACCGAGACCAACCAACCAAGACCGAGCTGCAG 2100  
Db ArgGlyArgGlnLysIleIleSerLeuThrGluThrAsnGlnLysAlaGluLeuGln 1059  
2101 GCATCCAGCTGGCGCTGCGAGCAGCGCAGCGAGGTGAACATCGTGCAGCGACCGCAG 2160  
Db AlaIleGlnLeuAlaLeuGlnAspSerGlySerGluValAsnValThrAspSerGln 1079  
2161 TAGCGCTGGCGCATCTCCAGCGCCCGCAGCAGCGAGCGAGCTGGTGAACCGAG 2220  
Db TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnLys 1099  
2221 ATCATCGAGCAGCTGATCAAGAAGGAGAGGTGTACCTGAGCTGGTGGCGCGCCACAG 2280  
Db IleIleGluGlnLeuIleLysLysGluArgValTyrLeuSerTrpValProAlaHisLys 1119  
2281 GGATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGCATCCGCAAGTGTG 2340  
Db GlyIleGlyGlnGluGlnValAspLysLeuValSerSerGlyIleArgLysValLeu 1139  
2341 TTCCTGGACGCGCATCGAT 2358  
Db PheLeuAspGlyIleAsp 1145

RESULT 15  
Q9WF79 PRELIMINARY; PRT; 1425 AA.  
ID Q9WF79 AC Q9WF79;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Gag-pol polyprotein.  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
NCBI\_TaxID=11676;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=99214383; PubMed=10196340;  
RA Novitsky V.A., Montano M.A., McLane M.F., Renjifo B., Vannberg F.,  
RA Foley B.T., Ndung'u T.P., Rahman M., Makhema M.J., Marlink R.,  
RA Essex M.;  
RT "Molecular cloning and phylogenetic analysis of human immunodeficiency  
RT virus type 1 subtype C: a set of 23 full-length clones from  
RT Botswana.";  
RL J. Virol. 73:4427-4432(1999).  
CC -1- SIMILARITY: Belongs to peptidase family A2.  
DR EMBL; AF1109569; AAD17087.1; --  
DR HSP; P24736; INCP.  
DR GO; GO:0019012; C:virion; IEA.  
DR GO; GO:0004190; F:aspartic-type endopeptidase activity; IEA.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008907; F:integrase activity; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0004523; F:ribonuclease H activity; IEA.  
DR GO; GO:0003723; F:RNA binding; IEA.  
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0005198; F:structural molecule activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0008270; F:zinc ion binding; IEA.  
DR GO; GO:0015074; P:DNA integration; IEA.  
DR GO; GO:0006310; P:DNA recombination; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.



```
||||| 838 ValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyValLysValArgGlnLeuCys 857
||||| 1501 AAGTGTCTGGCGCGCGCAAGCGCTGACGACATCGTCCCTGACCGAGGAGCGCGAG 1560
||||| 858 LysLeuLeuArgGlyAlaLysAlaLeuThrAspIleValProThrGluGluAlaGlu 877
||||| 1561 CTGGAGTGGCCGAGAACCGGAGATCTCGCGAGCCCGTGACGCGGTGTACTACGAC 1620
||||| 878 LeuGluLeuAlaGluAsnArgGluIleLeuLysGluProValHisGlyValTyrAsp 897
||||| 1621 CCCAGCAAGGACCTGTGGCGCGAGATCCAGAGAGCGGCCACGACGAGTGGACCTACGAC 1680
||||| 898 ProSerLysAspLeuIleAlaGluIleGlnLysGlnGlyAspGlnTrpThrTyrGln 917
||||| 1681 ATCTACGAGGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCC 1740
||||| 918 IleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaLysMetArgThrAla 937
||||| 1741 CACACCAACGACGTGAAGCAGCTGACGAGCGCGTGCAGAAAGATCGCCATGGAGAGCATC 1800
||||| 938 HisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleSerMetGluSerIle 957
||||| 1801 GTGATCTGGGCAAGACCCCAAGTTCGCGCTGCGCATCCAGAAAGAGACCTGGGAGACC 1860
||||| 958 ValIleTrpGlyLysIleProLysPheArgLeuProIleGlnLysGluAlaTrpGluAla 977
||||| 1861 TGTGGACCGACTACTCGCAGCGCCACCTGGATCCCGAGTGGGAGTTCTGTGACACCCCC 1920
||||| 978 TrpTrpThrAspTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrPro 997
||||| 1921 CCCCTGTGAAGCTGTGTACTGACGCTGAGAGAGGCCCATCATCGCGCGCGAGACCTTC 1980
||||| 998 ProLeuValLysLeuTrpTyrGlnLeuGluThrGluProMetAlaGlyAlaGluThrPhe 1017
||||| 1981 TACGTGACCGCGCGCCGACCGCGAGACCAAGATCCGGCAAGCGCGGCTACGTGACCGAC 2040
||||| 1018 TyrValAspGlyAlaAlaAsnArgGluThrLysIleGlyLysAlaGlyTyrValThrAsp 1037
||||| 2041 CGGGCGCGGAGAGATCGTGAGCGCTGACCGAGACCAACCAAGAGAGAGAGAGAGAGAG 2100
||||| 1038 LysGlyArgGlnGluValValThrLeuThrGluThrThrAsnGlnLysAlaGluLeuGln 1057
||||| 2101 GCCATCCAGCTGGCCCTGCAGGACGCGGAGCGAGCGAGGTGAACATCGTACCGCAGAGCCAG 2160
||||| 1058 AlaIleGlnLeuAlaLeuGlnAspSerGlyProGluValAsnIleValThrAspSerGln 1077
||||| 2161 TACGCCCTGGGCAATCATCCAGGCCCGACCGCGAGCAAGAGCGAGAGCGAGCTGGTGAACAG 2220
||||| 1078 TyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSerGluSerGluLeuValAsnGln 1097
||||| 2221 ATCATCGAGCAGCTGATCAGAGAGAGAGAGTGTACTGAGTGGTGGTGGCGGCCACAG 2280
||||| 1098 IleIleGluGlnLeuIleLysLysGluLysValTyrLeuSerTrpValProAlaHisLys 1117
||||| 2281 GGCATCGGCGCAACGAGCAGATCCACAAGCTGTGTGAGCAAGGGCATCCGCAAGGTGCTG 2340
||||| 1118 GlyIleGlyAsnGluLysValAspLysLeuValSerSerGlyIleArgGluValLeu 1137
||||| 2341 TTCCTGACCGCATCGAT 2358
||||| 1138 PheLeuAspGlyIleAsp 1143
```

Search completed: June 2, 2005, 05:05:45  
Job time : 305.158 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 2, 2005, 04:23:54 ; Search time 152.794 Seconds

(without alignments)  
11117.427 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 4555

Sequence: 1 gtcagcgcaccatggccga.....gggctagcaccggtgaattc 2457

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1465611 seqs, 345679903 residues

Total number of hits satisfying chosen parameters: 2931222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlh  
-Q=/cgn2\_1/USFO\_spool/US09610313/runat\_31052005\_155138\_15193/app\_query.fasta\_1.7893  
-DB=Published Applications\_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-TRANSPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blowsum62  
-THR MIN=0 -ALIGN=15 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-MAXLEN=2000000000 -USER=US09610313 @CGN 1.1 550 @runat 31052005\_155138\_15193  
-NCPU=6 -ICPU=3 -NO MAP -LARGESQRY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THRSADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Published Applications AA:
- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pcp.\*
  - 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pcp.\*
  - 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pcp.\*
  - 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pcp.\*
  - 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pcp.\*
  - 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pcp.\*
  - 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pcp.\*
  - 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pcp.\*
  - 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pcp.\*
  - 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pcp.\*
  - 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pcp.\*
  - 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pcp.\*
  - 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pcp.\*
  - 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pcp.\*
  - 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pcp.\*
  - 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pcp.\*
  - 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pcp.\*
  - 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pcp.\*
  - 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pcp.\*
  - 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	3527	77.4	998	16	US-10-332-413-4	Sequence 4, Appli
2	3480.5	76.4	850	16	US-10-332-413-10	Sequence 10, Appl
3	3454	75.8	1003	14	US-10-283-847-17	Sequence 17, Appl
4	3432	75.3	1015	17	US-10-634-165-9	Sequence 9, Appli
5	3409	74.8	739	16	US-10-093-953A-26	Sequence 26, Appl
6	3409	74.8	739	16	US-10-093-953A-31	Sequence 31, Appl
7	3387	74.4	1003	16	US-10-325-468-23	Sequence 23, Appl
8	3385	74.3	1003	16	US-10-325-468-35	Sequence 35, Appl
9	3384	74.3	1003	16	US-10-325-468-9	Sequence 9, Appli
10	3384	74.3	1003	16	US-10-325-468-20	Sequence 20, Appl
11	3376	74.1	995	15	US-10-296-734-2	Sequence 2, Appli
12	3362	73.8	995	15	US-10-296-734-1470	Sequence 1470, Ap
13	3208.5	70.4	1006	15	US-10-296-734-1471	Sequence 6, Appli
14	3140.5	68.9	1014	14	US-10-301-661A-6	Sequence 35, Appl
15	3076.5	67.5	1350	10	US-09-952-060-35	Sequence 35, Appl
16	3076.5	67.5	1350	16	US-10-380-641-35	Sequence 35, Appl
17	3076.5	67.5	1350	17	US-10-636-730-35	Sequence 6, Appli
18	2779.5	61.0	875	10	US-09-952-060-6	Sequence 6, Appli
19	2779.5	61.0	875	15	US-10-168-217A-6	Sequence 6, Appli
20	2779.5	61.0	875	16	US-10-380-641-6	Sequence 6, Appli
21	2779.5	61.0	875	17	US-10-636-730-6	Sequence 2, Appli
22	2774	60.9	850	10	US-09-952-060-2	Sequence 2, Appli
23	2774	60.9	850	15	US-10-168-217A-2	Sequence 2, Appli
24	2774	60.9	850	16	US-10-380-641-2	Sequence 2, Appli
25	2774	60.9	850	17	US-10-636-730-2	Sequence 8, Appli
26	2755.5	60.5	875	15	US-09-952-060-8	Sequence 8, Appli
27	2755.5	60.5	875	15	US-10-168-217A-8	Sequence 8, Appli
28	2755.5	60.5	875	16	US-10-380-641-8	Sequence 8, Appli
29	2755.5	60.5	850	17	US-10-636-730-8	Sequence 8, Appli
30	2750	60.4	850	10	US-09-952-060-4	Sequence 4, Appli
31	2750	60.4	850	15	US-10-168-217A-4	Sequence 4, Appli
32	2750	60.4	850	16	US-10-380-641-4	Sequence 4, Appli
33	2750	60.4	850	17	US-10-636-730-4	Sequence 4, Appli
34	2746	60.3	562	9	US-09-735-487-14	Sequence 14, Appl
35	2746	60.3	562	14	US-10-342-188-14	Sequence 2, Appli
36	2742	60.2	560	15	US-10-399-920-2	Sequence 87, Appl
37	2740	60.2	561	14	US-10-059-271-87	Sequence 10, Appl
38	2734	60.0	562	14	US-10-102-622-10	Sequence 1, Appli
39	2715	59.6	560	14	US-10-205-641-1	Sequence 12, Appl
40	2705.5	59.4	561	14	US-10-102-622-12	Sequence 1, Appli
41	2693	59.1	560	11	US-09-725-652-1	Sequence 29, Appl
42	2659	58.4	546	8	US-08-808-031A-29	Sequence 118, App
43	2646	58.1	979	14	US-10-271-181B-118	Sequence 88, Appl
44	2553	56.0	851	14	US-10-059-271-88	Sequence 3, Appli
45	2528	55.5	999	15	US-10-346-000A-3	

ALIGNMENTS

RESULT 1  
US-10-332-413-4  
; Sequence 4, Application US/10332413  
; Publication No. US20040116660A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnstone, Robert Edward  
; APPLICANT: Swannstrom, Ronald Ivar  
; APPLICANT: Morris, Lynn  
; APPLICANT: Karim, Salim Abdool  
; APPLICANT: Williamson, Carolyn  
; TITLE OF INVENTION: Process for the Selection of HIV-1 Subtype C Isolates, Their Genes nd Modifications and Derivatives T.  
; FILE REFERENCE: 45669-281993  
; CURRENT APPLICATION NUMBER: US/10/332,413  
; CURRENT FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: PCT/IB01/01208  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/216,995  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: ZA 2000/3437  
; PRIOR FILING DATE: 2000-07-10  
; PRIOR APPLICATION NUMBER: ZA 2000/4924

```
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 998
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-10-332-413-4

Alignment Scores:
Pred. No.: 1,148-163 Length: 998
Score: 3527.00 Matches: 662
Percent Similarity: 96.51% Conservative: 29
Best Local Similarity: 92.46% Mismatches: 21
Query Match: 77.43% Indels: 4
DB: 16 Gaps: 2

US-09-610-313B-32 (1-2457) x US-10-332-413-4 (1-998)
QY 223 TTCCGCGAGGACCTGGCTTCCCGGAGGCGCGAGTTCCTCCGAGCGAGCAAC 282
DB 1 PheArgGluAsnLeuAlaPheProGlnGlyGluAlaArgGluPheProSerGluGlnThr 20
QY 283 CCGCGCAACAGCCGAGCGCGAGCTGCAGGTGCGCGGCAACACCCCGCGAGCGAG 342
DB 21 ArgAlaAsnSerProThrSerArgGluLeuGlnValArgAsnAsnProArgSerGlu 40
QY 343 GCCGCGCGCGAGCGCGAGCGGACCTGAACTTCCCGAGATACACCTGTGGAGCGCCCC 402
DB 41 ThrGlyAlaGluArgGlyGlyThrLeuAsnPheProGlnIleThrLeuTrpGlnArgPro 60
QY 403 CTGTGTGACATCAAGTGGCGGCGGCGAGATCAAGGAGGCGCTGCTGGACACCGGCGCGAC 462
DB 61 LeuValSerIleLysIleGlyGlnThrArgGluAlaLeuLeuAsnThrGlyAlaAsp 80
QY 463 GACACCGTCTGGAGGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGC 522
DB 81 AspThrValLeuGluAspIleAsnLeuProGlyLysTrpLysProLysMetIleGly 100
QY 523 ATCGCGGGCTTCATCAGGTGGCGGCGAGTACGACGAGATCTGTGATCGAGATCTGGCGAAG 582
DB 101 IleGlyGlyPheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyLys 120
QY 583 AAGGCCATCGGACCGTGTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 642
DB 121 LysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnMet 140
QY 643 CTGACCCGAGCTGGGCTGCACCCCTGAACCTTCCCGATCAGCGCGCGGCGGCGGCGGCGG 702
DB 141 LeuThrGlnLeuGlyCysThrLeuAsnPheProIleSerProIleGluThrValProVal 160
QY 703 AAGCTGAAGCCGCGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 762
DB 161 LysLeuLysProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLys 180
QY 763 ATCAGGCGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 822
DB 181 IleLysAlaLeuThrAlaIleCysGluGluMetGluLysGluGlyLysIleThrLysIle 200
QY 823 GGCCCGGAGAACCCCTACAACACCCCGTGTTCGCGCATCAAGAAAGAGGAGGAGGAGGAGG 882
DB 201 GlyProGluAsnProTyAsnThrProIlePheAlaIleLysLysLysAspSerThrLys 220
QY 883 TGGCGCAAGCTGGTGGATTCCTCCGAGCTGAACAAAGCGGCGGCGGCGGCGGCGGCGG 942
DB 221 TrpArgLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluVal 240
QY 943 CAGCTGGGCGATCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1002
DB 241 GlnLeuGlyIleProHieProAlaGlyLeuLysLysLysLysValThrValLeuAsp 260
QY 1003 GTGGCGGAGCGCTACTTTCAGCGTGCCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1062
```

```
DB 261 ValGlyAspAlaTyPheSerValProLeuAspGluGlyPheArgLysTyThrAlaPhe 280
QY 1063 ACCATCCCGGAGCATCAACAGAGACCCCGGCGATCCGCTACCGAGTCAACAGTGTCTGCC 1122
DB 281 ThrIleProSerIleAsnAsnGluThrProGlyIleArgTyGlnTyAsnValLeuPro 300
QY 1123 CAGGCTGGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1182
DB 301 GlnGlyTrpLysGlySerProAlaIlePheGlnGlySerMetThrLysIleLeuGluPro 320
QY 1183 TTCCGCGCGCGCAACCCCGAGATCGTATCTACAG-----GCCCGCTCTAGCTGGCG 1236
DB 321 PheArgAlaGlnAsnProGluIleValIleTyGlnTyMetAspAspLeuTyValGly 340
QY 1237 AGCGACCTGGAGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1296
DB 341 SerAspLeuGluIleGlyGlnHisArgAlaLysIleGluGluLeuArgGluHisLeuLeu 360
QY 1297 CGCTGGGCGCTTCAACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1350
DB 361 LysTrpGlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMet 380
QY 1351 CCCATCGAGCTGCACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1410
DB 381 GlyTyGlnLeuHisProAspLysTrpThrValGlnProIleGlnLeuProGluLysAsp 400
QY 1411 AGCTGGACCGTGAACGACATCCAGAGCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1470
DB 401 SerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIle 420
QY 1471 TACCCGCGCATCAAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1530
DB 421 TyrProGlyIleLysValArgGlnLeuLysLysLeuArgGlyThrLysAlaLeuThr 440
QY 1531 GACATCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1590
DB 441 AspIleValProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeu 460
QY 1591 CGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1650
DB 461 LysGluProValHisGlyValTyTyAspProSerLysAspLeuIleAlaGluIleGln 480
QY 1651 AAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1710
DB 481 LysGlnGlyAspAspGlnTrpThrTyGlnIleTyGlnLeuProPhePheLysAsnLeuLys 500
QY 1711 ACCGCGCAAGTACGCGGAGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1770
DB 501 ThrGlyLysTyAlaLysArgThrThrHisThrAsnAspValLysGlnLeuThrGlu 520
QY 1771 GCGGTGCGAAGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1830
DB 521 AlaValGlnLysIleSerLeuGluSerIleValIleTrpGlyLysThrProLysPheArg 540
QY 1831 CTGCGCATCAAGAGGAGACCTGGGAGACCTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1890
DB 541 LeuProIleGlnLysGluThrTrpGluIleTrpIleTrpIleTrpIleTrpIleTrpIle 560
QY 1891 ATCCCGGAGTGGGAGTTCGTGAACACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1950
DB 561 IleProGluTrpGluPheValAsnThrProProLeuValLysLeuTyTrpTyGlnLeuGlu 580
QY 1951 AAGAGCGCATCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2010
DB 581 LysGluProIleAlaGlyAlaGluThrPheTyValAspGlyAlaAlaAsnArgGluThr 600
QY 2011 AAGATCGGCAAGGCGGCTACGTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2070
DB 601 LysIleGlyLysAlaGlyTyValThrAspArgGlyArgGlnLysIleValThrLeuSer 620
QY 2071 GAGACCCAAACAGAGACCGAGCTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2130
DB 621 GluThrThrAsnGlnLysThrGluLeuGlnAlaIleGlnLeuAlaLeuGlnAspSerGlu 640
```









```
QY 268 -----CCAGCGAGCAGAACCGCGCCAAACAGACCCCAACGAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrArgAlaAsnSerProThrArg 40
QY 304 CCGAGCTCAGGTG-----CGCGCGCAACACCCCGCAGCGAGCGCGCGCCAGGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCGACCTG-----NACTTCCCAGATACCGTGTGCGAGGCGCCCTCGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGTGGCGCGCAGATCAAGAGGCGCTGTGGACACCGCGCGCGCAGCACCGGTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTGCCCGCAAGTGGAAAGCCCAAGATATCGGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyIleGlyGly 120
QY 532 TTCATCAAGTGGCGCGCAGTACACACAGATCCTGATCGAGATCTGGGCAAGAGCCATC 591
Db 121 PheIleLysValArgGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle 140
QY 592 GGCACCGTCTGATCGGCGCCACCCCGTGAACATCATCGCGCGCGCAACATGTCACCCAG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGTGCACCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGAAGTGAAG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeuLys 180
QY 712 CCGCGCATGGAGCGGCCCAAGTGAAGCATGGCCCTGACCGAGAGAAGATCAAGGCC 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCATCTGGAGGAGATGAGAGAGGCGGCAGATCACCAAGATCGGCCCGAG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
QY 832 AACCCCTACAACACCCCGCTGTTCCGCATCAAGAAGAAGCAGACCAAGTGGCGCAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGGTGAATCTCCGAGCTGAACAGCGCAACCGAGACTTCTGGAGAGTGCAGTGGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCGCCCGCGCGCTGAAGAGAGAGAGCGGTGACCGTCTGAGCTGGCGGCAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTCAGCGTGCCTGACAGAGACTTCCGCAAGTACACCGCTTACCATCCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACAACAGACACCCCGCATCCGCTACCAAGTACAAGTCTGCCCGCAGGCTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGCGAGCCCGAGCATCTTCAGACAGCATGACCAAGATCTCGAGGCCCTTCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCCGAGATCGTGATCTACACAG-----GCCCGCTGTAGTGGCGAGCACCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeu 360
QY 1246 GAGATCGGCGACCGCGCCCAAGATCGAGGAGCTGCGCAAGCACCTGCTGCGCTGGGCG 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly 380
```

RESULT 5

```
QY 1306 TTCACCAACCCCGCAACAAGACACAGAAAGAGAGCCCGCTTCTCTG-----CCCATCGAG 1359
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1360 CTGCAACCCCGCAACAAGTGGACCGTGCAGCCCATCGAGTGTCCCGAAGAGAGAGCTGGACC 1419
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1420 GTGAACGACATCCAGAGCTGGTGGGCAAGCTGAAGTGGGCGCAGCAGATCTACCCCGC 1479
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGly 440
QY 1480 ATCAAGTGGCCCGCAGCTGTCAAGCTGTCCGCGCGCGCAAGGCCCTCACGACATCTGTG 1539
Db 441 IleLysValArgGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1540 CCCTGTACCCGAGGAGCGAGCTGGAGTGCCTGCAGAACCCGCGAGATCCTGCGGAGGCC 1599
Db 461 ProLeuThrGluGluAlaGluLeuAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1600 GTGCACGCGGTGTACTAGCACCCAGCAAGACCTGGTGGCGGAGATCCAGACGAGGCC 1659
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1660 CACGACCATGTGCACCTACCATGATCTACCAGAGCGCTTCAAGAACCTGAAGACCGGCAAG 1719
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1720 TACGCCAAGATGCGCACCGCCCACACCAACGACGTGAAGCAGCTGACCGAGCGCGTGCAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGln 540
QY 1780 AAGATCGCCATGGAGAGATCGTATCTGGGGCAAGACCCCAAGTTCGCGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1840 CAGAAGGAGACTCGGAGACCTGTGGACCTGACACTACTGCGCAGGCCACCTGATCCCGCAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1900 TGGAGATTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTACCACTGGAGAAGAGGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1960 ATCATCGCGCGCGAGACTTCTACGTGACCGCGCGCGCAACCCGCGAGACCAAGATCCGC 2019
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCCGCGCTACGTGACCGACCGCGCGCGCAGAGAGATCGTGAGCTGACCGAGACCAACC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACCAGAAAGACCGAGCTGCAGGCCATCCAGCTGGCGCTGACGAGCAGCGCGCAGGAGTG 2139
Db 641 AsnGlnLysThrGluLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCGTGAACGACGACGAGTACGCCCTGGGCGATCATCCAGGCCCGCGCAGCAAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGAGCGAGCTGGTGAACCAACCATCATCATGAGCAGCTGATCAAGAAGGAGAGAGTGTACCTG 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCGCGCCCAACAGGGCATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSer 720
QY 2320 AAGGCGATCCGCAAGGTGCTGTTCTGAGCGCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

US-10-093-953A-26

; Sequence 26, Application US/10093953A

; Publication No. US20040105871A1

; GENERAL INFORMATION:

; APPLICANT: Robinson, Harriet L.

; APPLICANT: Smith, James M.

; APPLICANT: Hua, Jian

; APPLICANT: Moss, Bernard

; APPLICANT: Amara, Rama

; APPLICANT: Wyatt, Linda

; APPLICANT: Earl, Patricia

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING

; FILE OF INVENTION: AN IMMUNE RESPONSE

; FILE REFERENCE: 12804-005002

; CURRENT APPLICATION NUMBER: US/10/093.953A

; CURRENT FILING DATE: 2002-03-08

; PRIOR FILING DATE: US 60/186,364

; PRIOR FILING DATE: 2000-03-02

; PRIOR FILING DATE: US 60/251,083

; PRIOR FILING DATE: 2000-12-01

; PRIOR APPLICATION NUMBER: US 09/798,675

; PRIOR APPLICATION NUMBER: PCT/US01/06795

; PRIOR APPLICATION NUMBER: US 60/324,845

; PRIOR FILING DATE: 2001-09-25

; PRIOR APPLICATION NUMBER: US 60/325,004

; PRIOR FILING DATE: 2001-09-26

; NUMBER OF SEQ ID NOS: 52

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 26

; LENGTH: 739

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: protein encoded by construct of vaccine vector

; OTHER INFORMATION: pGA2 and insert JS2 expressing clade HIV-1 VL

US-10-093-953A-26

Alignment Scores:

Pred. No.:	6.15e-158	Length:	739
Score:	3409.00	Matches:	647
Percent Similarity:	92.91%	Conservative:	34
Best Local Similarity:	88.27%	Mismatches:	32
Query Match:	74.84%	Indels:	20
DB:	16	Gaps:	5

US-09-610-313B-32 (1-2457) x US-10-093-953A-26 (1-739)

Qy	220	TTCTTCGCGAGGACCTGGCTTCCCGCAGGCAAGCCCGCGAGTTCCCGCAGCGAGCAG	279
Db	1	PhePheargGluaspLeuAlaPheLeuGlnGlyLysAlaargGluPheSerSerGluGln	20
Qy	280	AACGCG-----GCAACAGCCCGCCAGC	303
Db	21	ThrArgAlaasnSerProThrIleSerSerGluGlnThrGlyAlaasnSerProThrArg	40
Qy	304	CGCGAGCTGCAGTG-----CGCGGCAACCCCGCAGCGAGCCGCGCGCGAGCGC	357
Db	41	ArgGluLeuGlnValTrpGlyArgaspAsnAsnSerProSerGluAlaGlyAlaaspArg	60
Qy	358	CAGGCGACCCCTG-----AACTTCCCGCAGATCACCTGTGGCAGCGCCCGCTGGTGAGC	411
Db	61	GlnGlyThrValSerPheasnPheProGlnIleThrLeuTrpGlnArgProLeuValThr	80
Qy	412	ATCAAGTGGCGGCCAGATCAAGAGCGCTGTGACACCGCGCGCGCGAGCACCGTG	471
Db	81	IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspThrVal	100
Qy	472	CTGGAGGAGATGACCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGC	531
Db	101	LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyGlyIleGlyGly	120

Qy	532	TTCAATCAAGTGGCGCGAGTACGACCATCTGATCGAGATCTCGGCAAGAGCGCCATC	591
Db	121	PheIleLysValargGlnTrpAspGlnIleLeuIleGluIleCysGlyHisLysAlaIle	140
Qy	592	GGCAGCGTGTGATCGCGCCCGCCCGTGAACATCATCGCGCGCACATGCTGACCCAG	651
Db	141	GlyThrValLeuValGlyProThrProValasnIlelleGlyArgasnLeuLeuThrGln	160
Qy	652	CTGGGCTGCACCCCTGAACCTTCCCATCAGCCCGCATCGAGACCGCTGCCGTGAAGCT	711
Db	161	IleGlyCysThrLeuasnPheProIleSerProIleGluThrValProValLysLeuLys	180
Qy	712	CCGCGCATGACGCGCCCGCCCGTGAACAGTGAAGTGGCCCTGACCGAGGAGAGATCAAG	771
Db	181	ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla	200
Qy	772	CTGACCGCCCATCTGCGAGGAGATGAGAGAGGCGCAAGATCACCAAGATCGCGCCG	831
Db	201	LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu	220
Qy	832	AAGCCCTTACACACCCCGCTGTTCGCCCATCAAGAGAGGACAGCACCAAGTGGCGCA	891
Db	221	AsnProTyraasnThrProValPheAlaIleLysLysLysaspSerThrLysTrpArgLys	240
Qy	892	CTGGTGGACTTCCCGAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGCAGCTGGC	951
Db	241	LeuValAspPheargGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly	260
Qy	952	ATCCCCCAGCCCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1011
Db	261	IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp	280
Qy	1012	GCCTACTTTCAGCGTCCCGTGGACGAGGACTTCCGCAAGTACACCGCTTCCACCATCCC	1071
Db	281	AlatyPheSerValProLeuAspGluaspPheargLysTrpThrAlaPheThrIlePro	300
Qy	1072	AGCATCAACAAACGAGACCCCGCGCATCCGTTACAGTACAACTGCTGCCCGAGGCTGG	1131
Db	301	SerIleAsnAsnGluThrProGlyIleArgTyrglnTrpAsnValLeuProGlnGlyTrp	320
Qy	1132	AAGGCGAGCCCGAGCATCTTCCAGAGGAGCATGACCAAGATCTCGGAGCCCTTCCGCG	1191
Db	321	LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys	340
Qy	1192	CGCAACCCCGAGATCGTGTACTTACCAG-----GCCCGCTGTACGTGGGCGAGCGCTG	1245
Db	341	GlnasnProaspIleValIleTyrglnTyrglnMetAsnAspLeuTyrglnLysSeraspLeu	360
Qy	1246	GAGATCGCGCAGCACCGCGCCAAAGATCGAGAGCTGCGCAAGCACTGCTGCTGGGCG	1305
Db	361	GluIleGlyGlnHisargThrLysIleGluGluLeuArgGlnHisLeuLeuArgTrpGly	380
Qy	1306	TTCAACACCCCGCAGAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1359
Db	381	LeuThrThrProaspLysLysHisGlnLysGluProPheLeuTrpMetGlyTyrglu	400
Qy	1360	CTGCACCCCGCAGCAAGTGGAGCGCTGCAGCTGCAGCTGCCGAGAGAGAGAGTGGACC	1419
Db	401	LeuHisProaspLysTrpThrValGlnProIleValLeuProGluLysaspSerTrpThr	420
Qy	1420	GTGAACGACATCCAGAGCTGGTGGGCAAGCTGAACCTGGCGCCAGCCAGATCTACCCCG	1479
Db	421	ValasnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrgly	440
Qy	1480	ATCAAGTGGCGCGCAGCTGTCAGAGCTGCTCGCGCGCGCCAGGCGCTGACCGACATCGT	1539
Db	441	IleLysValargGlnLeuCysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle	460
Qy	1540	CCCTGACCGAGGAGCGCGAGCTGGAGCTGCGCGAGAACCGCGAGATCTCTGCGCGAGCC	1599
Db	461	ProLeuThrGluGluAlaGluLeuGluLeuAlaGluLeuAsnArgGluIleLeuLysGluPro	480
Qy	1600	GTGCACGCGCTGTACTACGACCCCGCAGCAGGAGCTGTGTGGCGCGAGATCCAGAGCG	1659

```
Db 481 ValHisGlyValTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1660 CAGACCACTGAGTACATACAGATCTACAGAGAGCCCTTCAAGAACTGAAGACCGGCAAG 1719
Db 501 GlnGlyGlnTrpThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1720 TACGCCAAGATCGCACCGCCACACCAACAGCGTGAAGCAGCTCACCGAGCGCGTGCAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
QY 1780 AAGATCGCCATCGAGAGATCTGTATCTGGGCAAGACCCCAAGTTCGGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1840 CAGAAGAGACCTGGAGACCTGTGGACCGACTACTGGACGAGCCACCTGATGCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpThrTrpThrGluTyrTrpGlnAlaThrTrpIleProGlu 580
QY 1900 TGGGAGTTCTGTAACACCGCCCTCTGTGAGCTGTGTACCATCTGGAGAGAGCGCC 1959
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1960 ATCATCGGCGCGAGACCTTCTACCTGACCGCGCGCCCAACCGCGAGACCAAGATCGCG 2019
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCGCGCTACTGATACCGACCGCGCGCGGAGAGATCTGTGAGCTGACCGAGACCGCC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACCAAGACCGACCTGACCGCCATCCAGCTGGCGCCCTCGAGGACGCGGACGAGCTG 2139
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuLeuAspSerGlyLeuGluVal 660
QY 2140 AACATCGTACCGACGACGCGGATCGCTTGGCGCATCATCCAGCGCCAGCCGACAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGACGAGCTGGTGAACAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGT 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTCCCGCCCAAGGGCATCGCGGCAAGCAGCAGATCGACAGAGCTGTGTGAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSer 720
QY 2320 AAGGGCATCCGCAAGTGTCTTCTGACGCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733
```

## RESULT 6

```
US-10-093-953A-31
; Sequence 31, Application US/10093953A
; Publication No. US20040105871A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Harriet L.
; APPLICANT: Smith, James M.
; APPLICANT: Hua, Jian
; APPLICANT: Moss, Bernard
; APPLICANT: Amara, Rama
; APPLICANT: Wyatt, Linda
; APPLICANT: Earl, Patricia
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR GENERATING
; FILE REFERENCE: 12804-005002
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/186,364
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/251,083
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/798,675
```

```

; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/US01/06795
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: US 60/324,845
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/325,004
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: protein encoded by construct of vaccine vector
; OTHER INFORMATION: pCAL and vaccine insert expressing clade B HIV-1
; OTHER INFORMATION: Gag-Pol
US-10-093-953A-31

Alignment Scores:
Pred. No.: 6,15e-158 Length: 739
Score: 3409.00 Matches: 647
Percent Similarity: 92.91% Conservative: 34
Best Local Similarity: 88.27% Mismatches: 32
Query Match: 74.84% Indels: 20
DB: 16 Gaps: 5

US-09-610-313B-32 (1-2457) x US-10-093-953A-31 (1-739)
QY 220 TTCTTCCCGAGGACCTGGCTTCCCTCCCGAGGCAAGGCCCGAGTTCCCGAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheLeuGlnGlyLysAlaArgGluPheSerSerGluGln 20
QY 280 AACGCG-----GCCAAGCCCCCACCAGC 303
Db 21 ThrArgAlaAsnSerProThrIleSerSerGluGlnThrGlyAlaAsnSerProThrArg 40
QY 304 CGCGAGCTGCAGGTG-----CGCGGCGACAAACCCCGCAGCGCGCGCGCGCGCGCGC 357
Db 41 ArgGluLeuGlnValTrpGlyArgAspAsnAsnSerProSerGluAlaGlyAlaAspArg 60
QY 358 CAGGCAACCTCG-----AACTTCCCGCAGATCACCTGTGTGCGAGCGCCCTGTGTGAGC 411
Db 61 GlnGlyThrValSerPheAsnPheProGlnIleThrLeuTrpGlnArgProLeuValThr 80
QY 412 ATCAAGTGGCGCGCGCAGATCAAGAGCGCTCTGTGGACACCGCGCGCGCGCGAGACCGCTG 471
Db 81 IleLysIleGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThrVal 100
QY 472 CTGGAGGAGATGAGCTCGCGCGCAAGTGGAGCCCAAGATGATCGCGCGCATCGCGCGC 531
Db 101 LeuGluGluMetSerLeuProGlyArgTrpLysProLysMetIleGlyLysIleGly 120
QY 532 TTCATCAAGTGGCGCGCAGTACGACACAGATCTCTGTATCGAGATCTGCGCAAGAGCCCATC 591
Db 121 PheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyHisLysAla 140
QY 592 GGCACCGTGTGATCGCGCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
Db 141 GlyThrValLeuValGlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGln 160
QY 652 CTGGGCTGCACCTGAACTTCCCATCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 711
Db 161 IleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeu 180
QY 712 CCCGCATGGAGCGGCCCAAGGTGAAGCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 771
Db 181 ProGlyMetAspGlyProLysValLysGlnTrpProLeuThrGluGluLysIleLysAla 200
QY 772 CTGACCGCGCTCTGGAGGAGATGAGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 831
Db 201 LeuValGluIleCysThrGluMetGluLysGluGlyLysIleSerLysIleGlyProGlu 220
```

```
QY 832 AACCCCTACACACCCCGCTGTTCCCATCAAGAAAGAGACACCAAGTGGCGAAG 891
Db 221 AsnProTyrAsnThrProValPheAlaIleLysLysAspSerThrLysTrpArgLys 240
QY 892 CTGTGTGACTTCGCGAGCTGAACAACGACGACCCAGGACTTCGGGAGGTGCAGCTGGC 951
Db 241 LeuValAspPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGly 260
QY 952 ATCCCCCACCACCGCCCTGAAGAAGAAGAGCGTGCACCGTGTGGCGGAC 1011
Db 261 IleProHisProAlaGlyLeuLysLysLysSerValThrValLeuAspValGlyAsp 280
QY 1012 GCCTACTTCAGCTGCGCCCTGGAGAGACTTCGCGAAGTACACCGCTTCACCATCCC 1071
Db 281 AlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIlePro 300
QY 1072 AGCATCAACACGAGACCCCGGCATCCGCTACCAAGTACACGTCGTCGCCCGAGGTGG 1131
Db 301 SerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrp 320
QY 1132 AAGGCGACGCCCGCATCTTCCAGAGCAGCATGACCAAGATCCTGGAGCCCTTCCGCGCC 1191
Db 321 LysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheLysLys 340
QY 1192 CGCAACCCGAGATCGTGATCTACAG-----GCCCCCTGTAGTGGGAGGACCTG 1245
Db 341 GlnAsnProAspIleValIleTyrGlnTyrMetAsnAspLeuTyrValGlySerAspLeu 360
QY 1246 GAGATCGCGCAGCAGCCGCGCCAGATCGAGGAGCTGCGCAGCAGCTGCGCTGGCGGC 1305
Db 361 GluIleGlyGlnHisArgThrLysIleGluLeuArgGlnHisLeuLeuA-gtrpGly 380
QY 1306 TTCACACCCCGCACAAGAAGCACCAGAAGAGCGCCCTTCTCTG-----CCCATCGAG 1359
Db 381 LeuThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGlu 400
QY 1360 CTGCACCCCGACAAGTGGACCGTGCAGCCCATCGAGCTGCCGAGAGAGAGAGCTGACCC 1419
Db 401 LeuHisProAspLysTrpThrValGlnProIleValLeuProGluLysAspSerTrpThr 420
QY 1420 GTGACACATCCAGAGCTGTGGGCAAGCTGAAGTGGGCCAGCAGATCACTCCCGGC 1479
Db 421 ValAsnAspIleGlnLysLeuValGlyLysLeuAsnThrAlaSerGlnIleTyrProGly 440
QY 1480 ATCAAGTGGCCAGCTGTGAAGCTGCTGCGCGCGCCAGGCGCTGACCATCGCTG 1539
Db 441 IleLysValArgGlnLysLysLysLeuLeuArgGlyThrLysAlaLeuThrGluValIle 460
QY 1540 CCCCTGACCGAGGAGCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTCTGCGCGAGCC 1599
Db 461 ProLeuThrGluGluAlaGluLeuGluAlaGluAsnArgGluIleLeuLysGluPro 480
QY 1600 GTGACCGCTGTACTAGCACCCAGACACCTGTGGCGGAGATCCAGACGAGGC 1659
Db 481 ValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGlnGly 500
QY 1660 CACGACCATGTGCCTACCATCATCTACAGAGCGCTTCAAGAACCTTGAAGACCGCAAG 1719
Db 501 GlnGlyGlnThrThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLys 520
QY 1720 TACGCCAAGATGCGACCGCCACACCAAGACGTGAAGCAGCTGACCGAGCGCGTGCAG 1779
Db 521 TyrAlaArgMetArgGlyAlaHisThrAsnAspValLysLeuLeuThrGluAlaValGln 540
QY 1780 AAGATCCGCATGGAGAGATCGTGATCTGGGGCAAGACCCCAAGTTCGCTGCCCATC 1839
Db 541 LysIleThrThrGluSerIleValIleTrpGlyLysThrProLysPheLysLeuProIle 560
QY 1840 CAGAAGAGACCTGGGAGACCTGTGGACCGACTACTGGGAGCGCCACTGTGATCCCGAG 1899
Db 561 GlnLysGluThrTrpGluThrTrpTrpTrpGlnIleAlaThrTrpIleProGlu 580
QY 1900 TGGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGTGACCATCGTGGAGAGGAGGCC 1959
```

```
Db 581 TrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluLysGluPro 600
QY 1960 ATCATCGCGCCCGAGACTTCTACGTGGAGCGGCGCCCAACCGCGAGACCAAGATCGGC 2019
Db 601 IleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGly 620
QY 2020 AAGCGCGCTACGTAGCCGACCGGCGCGGAGAGATCGTGCAGCTGACCGACCGACACC 2079
Db 621 LysAlaGlyTyrValThrAsnLysGlyArgGlnLysValValProLeuThrAsnThrThr 640
QY 2080 AACCAAGACCGAGCTGCAGGCGCATCCAGCTGCGCCCTGAGGACGACGCGGAGGAGTG 2139
Db 641 AsnGlnLysThrGlnLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluVal 660
QY 2140 AACATCGTGCACCGACAGCAGTACCGCTCGGCGCATCATCCAGGCGCCAGCCGCAAGAGC 2199
Db 661 AsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLysSer 680
QY 2200 GAGAGCGAGCTGTGAACAGATCATCGAGCAGCTGATCAAGAGGAGAGAGTGTACCTG 2259
Db 681 GluSerGluLeuValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeu 700
QY 2260 AGCTGGGTGCGCCCGCACAAGGCGCATCGCGCGCAACGAGCAGATCGACAGCTGTCGAGC 2319
Db 701 AlaTrpValProAlaHisLysGlyIleGlyAsnGlnGlnValAspLysLeuValSer 720
QY 2320 AAGGCGATCCGCAAGGTGCTGTTCTCGGCGCATCGAT 2358
Db 721 AlaGlyIleArgLysIleLeuPheLeuAspGlyIleAsp 733

RESULT 7
US-10-325-468-23
; Sequence 23, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; FILE REFERENCE: MODELS OF HIV-1 PATHOGENESIS
; CURRENT APPLICATION NUMBER: US/10/325,468
; PRIOR FILING DATE: 2002-12-19
; CURRENT APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.10 protein Pol
US-10-325-468-23

Alignment Scores:
Pred. No.: 7,38e-157 Length: 1003
Score: 3387.00 Matches: 634
Percent Similarity: 94.17% Conservative: 45
Best Local Similarity: 87.93% Mismatches: 34
Query Match: 74.36% Indels: 8
DB: 16 Gaps: 4

US-09-610-313B-32 (1-2457) x US-10-325-468-23 (1-1003)
QY 220 TTCTTCGCGAGACCTGGCTTCCCGAGGCAAGCGCCGAGTTCCTCCGAGGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20
QY 280 AACCGCGCCACACCGCCCGAGCTGAGCTGAGTG-----CGCGGCGACAACCCC 333
Db 21 ThrArgAlaAsnSerProIleArgArgGlnValTrpArgA-gaspAsnAsnSer 40
```





APPLICANT: Soong, Nay Wei  
APPLICANT: Pekrun, Katja  
APPLICANT: Shibata, Riri  
TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
FILE REFERENCE: 0166.210US  
CURRENT APPLICATION NUMBER: US/10/325,468  
CURRENT FILING DATE: 2002-12-19  
PRIOR APPLICATION NUMBER: US 60/343,524  
PRIOR FILING DATE: 2001-12-21  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35

LENGTH: 1003

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: recombinant / chimeric sequence: clone P8A26 protein Pol  
US-10-325-468-35

#### Alignment Scores:

Pred. No.: 9.23e-157 Length: 1003  
Score: 3385.00 Matches: 635  
Percent Similarity: 94.17% Conservatives: 44  
Best Local Similarity: 88.07% Mismatches: 34  
Query Match: 74.31% Indels: 8  
DB: 16 Gaps: 5

US-09-610-313B-32 (1-2457) x US-10-325-468-35 (1-1003)

QY	220	TTCTTCGCGAGGACCTGGCTTCCCGCAGGCGAGCCCGCGAGTTCCTCCAGCGAGCAG	279
DB	1	PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln	20
QY	280	AACGCGCCGACACGCCACCGCGAGCTGCGAGTGCAGTG---CGCGCGAC---AACCCC	333
DB	21	ThrArgAlaAsnSerProIleArgGluArgGlnValTrpArgGlyAspAsnSer	40
QY	334	CGCAGCGAGCGCGCGCGAGCGCGAGCGCGACCTG-----AACTTCCCGAGATCAC	387
DB	41	LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr	60
QY	388	CTGTGGCAGCGCCCTCGTGGATCAAGTGGCGGCGCAGATCAAGAGGCGCTGCTG	447
DB	61	LeuTrpGlnArgProLeuValThrIleLysIleGlyLysGluLeuLysGluAlaLeu	80
QY	448	GACACCGCGCGAGCACCGCTGCTGAGAGATGAGCTGCCCGGCAAGTGAAGCCC	507
DB	81	AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro	100
QY	508	AAGATGATCGCGGCATCGCGCTTCATCAAGTGGCGCAGTACGACGACGATCCTGATC	567
DB	101	LysMetIleGlyIleGlyIleGlyPheIleLysValArgGlnIleAspGlnIleProIle	120
QY	568	GAGATCTGCGGCAAGAGGCGATCGGACCGTGTGATCGCGCCCGCCCGCGTGAACATC	627
DB	121	AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle	140
QY	628	ATCGCGCGCAACATGCTGACCCAGCTGGTGTGACCTGACCTGACCTTCCCATGACCCCATC	687
DB	141	IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle	160
QY	688	GAGACCTGCGCTGAGCTGAAGCCCGGATGACCGCCCGCCCGAGGTGAAGCTGGCCC	747
DB	161	GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro	180
QY	748	CTGACCGGAGGAGATCAAGCGCTGACCGCCATCTCGCAGGAGAGATGAGAGAGGGGC	807
DB	181	LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly	200
QY	808	AAGATCAACCAAGATCGGCGCGAGACCCCTTACACACCCCGCTGTGTCATCAAGAAG	867
DB	201	LysIleSerLysIleGlyProGluAsnProIleAsnThrProValPheAlaIleLysLys	220

QY	868	AAGGACAGCACCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAACAGCGACCCAG	927
DB	221	LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln	240
QY	928	GACTTCTGGAGGTGAGCTGGGATCCCGACCCCGCGCGCTGAAGAAGAAAGAGC	987
DB	241	AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer	260
QY	988	GTGACCGTGTGGACGTGGCGCGACCTTCTAGCGTGCCTCGGAGGAGGACTTCGCG	1047
DB	261	ValThrValLeuAspValGlyAspAlaTrpPheSerValProLeuAspLysAspPheArg	280
QY	1048	AAGTACACCGCCTTCCACATCCCGACGATCAACACGAGACCCCGCGCATCCGCTACCG	1107
DB	281	LysTrpThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTrpGln	300
QY	1108	TACACGTGTGCTGCCCGCGAGGTGAGGCGGAGCCCGAGCATCTTCCAGAGCAGCATGAC	1167
DB	301	TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr	320
QY	1168	AAGATCTGTGAGCGCTTCCGCGCGCGCAACCCCGAGATCGTGATCTACCGAG-----GCC	1221
DB	321	LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleIleIleIleIleIleIle	340
QY	1222	CCCTGTGTACGTGGCGAGCGACCTGGAGATCGGCGAGCACCGCGCGCAAGATCGAGGAGCTG	1281
DB	341	AspLeuTrpValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluLeu	360
QY	1282	CGCAAGCCTGTGTGGCTGGGCTTCCACCGCCCGACAGAACGACCAAGAGAGAGCC	1341
DB	361	ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro	380
QY	1342	CCCTTCTCTG-----CCCATCGAGCTGCACCGCGCAAGTGGACCGTGCACCCATCGAG	1395
DB	381	ProPheLeuTrpMetGlyTrpGluLeuHisProAspLysTrpThrValGlnProIleVal	400
QY	1396	CTGCGCGAGAGTGCAGCGCTGAACGACATCCAGAACGCTGGTGGGCAAGCTCAAC	1455
DB	401	LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn	420
QY	1456	TGGCGCAGCGAGATCTACCCCGCGCATCAAGTGGCGCGAGTGTGCAAGTGTGCGCGC	1515
DB	421	TrpAlaSerGlnIleTrpAlaGlyIleLysValLysGlnLeuLysLysLeuArgGly	440
QY	1516	GCCAAGCGCTGACCGCATCGTGCCTTACCGAGGAGCGCGAGTGGAGCTGGCGCGAG	1575
DB	441	ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuLeuAlaGlu	460
QY	1576	AACCGCGAGATCTGCGCGAGCGCGTGCACGGCTGTACTACGACCCCGAGCAAGGACCTG	1635
DB	461	AsnArgGluLeuLysGluProValHisGlyValTrpTrpAspProSerLysAspLeu	480
QY	1636	GTGCGCGAGATCCAGAACGAGCGCGACGACGAGTGGACCTTACGAGATCTACGAGAGCC	1695
DB	481	IleValGluIleGlnLysGlnGlyGlnTrpThrTrpGlnIlePheGlnGluPro	500
QY	1696	TTCAAGACCTGAGACCGCGAGTAGCCCAAGATGGCGACCGCGCCCGCACCAAGAGCTG	1755
DB	501	PheLysAsnLeuLysThrGlyLysTrpAlaLysThrArgGlyAlaHisThrAsnAspVal	520
QY	1756	AAGCAGCTGACCGAGCGCGTGCAGAGATCGCATGGAGAGCATCGTGATCTGGGGCAAG	1815
DB	521	LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys	540
QY	1816	ACCCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGGGAGACCTGGTGCACCGACTAC	1875
DB	541	IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr	560
QY	1876	TGGCAGCGCACCTGAGTCCCGAGTGGGAGTTGTTGTAACACCCCGCTGTGGTGAAGCTG	1935
DB	561	TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu	580

QY 1936 TGTACCACTGGAGAGAGCCCATCATCGCGCGGACAGCCTTCTACGTGACGCGGCC 1995  
 Db 581 TptTyrGlnLeuGluLysGluProIleValGluAlaGluThrPheTyrValAspGlyAla 600  
 QY 1996 GCGACCGCGAGACCAAGATCGGCAAGCGCGGCTAGCTACGACCGGGCGCGGAGAAG 2055  
 Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620  
 QY 2056 ATCGTGAGCTCAGCGAGACCAACACCAAGAGACCGAGCTCAGGSCCATCCAGCTGGCC 2115  
 Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaHisLeuAla 640  
 QY 2116 CTCGAGGACGCGGAGCGAGGTGAACATCGTACCGACGAGCGAGTACGCCCTGGGCGATC 2175  
 Db 641 LeuGlnAspSerGlyLeuGluValAlaAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660  
 QY 2176 ATCCAGGCGCCAGCGACAAAGAGCGAGAGCGAGTGTGTGAACAGATCATCGAGCAGCTG 2235  
 Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680  
 QY 2236 ATCAAGAGGAGAGGTGTACCTGAGCTGGGTGCGCGCCGCCACAGGGGCATCGCGGCAAC 2295  
 Db 681 IleLysLysGluLysValTyrLeuThrTrpIleProAlaHisLysGlyIleGlyAsn 700  
 QY 2296 GAGCAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTGTCTCGACGCGCATC 2355  
 Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgArgValLeuPheLeuAspGlyIle 720  
 QY 2356 GAT 2358  
 Db 721 Glu 721

RESULT 9

US-10-325-468-9  
 ; Sequence 9, Application US/10325468  
 ; Publication No. US20040101823A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Soong, Nay Wei  
 ; APPLICANT: Pekrun, Katja  
 ; APPLICANT: Shibata, Riri  
 ; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL  
 ; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS  
 ; FILE REFERENCE: 0166.210US  
 ; CURRENT APPLICATION NUMBER: US/10/325,468  
 ; CURRENT FILING DATE: 2002-12-19  
 ; PRIOR APPLICATION NUMBER: US 60/343,524  
 ; PRIOR FILING DATE: 2001-12-21  
 ; NUMBER OF SEQ ID NOS: 56  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9  
 ; LENGTH: 1003  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.4 protein Pol;  
 ; OTHER INFORMATION: clone 1.26 protein Pol; clone P10.21 protein Pol;  
 ; OTHER INFORMATION: clone P10.26 protein Pol  
 US-10-325-468-9

Alignment Scores:

Align. No.: 1,03e-156 Length: 1003  
 Score: 3384.00 Matches: 634  
 Percent Similarity: 94.04% Conservatives: 44  
 Best Local Similarity: 87.93% Mismatches: 35  
 Query Match: 74.29% Indels: 8  
 DB: 16 Gaps: 4

US-09-610-313b-32 (1-2457) x US-10-325-468-9 (1-1003)

QY 220 TTCTTCGGAGAGACTCGCTGGCTTCCCGCAGGGCAGGCGCGCGAGTTCGCCAGCAGCAG 279  
 Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGluGln 20

QY 280 AACCGCGCAACAGCGCCACCGCGCGAGCTGCAGGTG-----CGCGGCGACAAACCCC 333  
 Db 21 ThrArgAlaAsnSerProIleArgArgGluArgGlnValTrpArgArgAspAsnSer 40  
 QY 334 CGCAGCGAGCGCGCGCGCGAGCGCGCGAGCGCGCGT-----AATCTCCCGAGATCACC 387  
 Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60  
 QY 388 CTGTGGCAGCGCCCGCTGCTGAGCATCAAGGTGGCGCGCCAGATCAAGAGGAGCGCTGCTG 447  
 Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80  
 QY 448 GACACCGCGCGCGCGAGCACACCGCTGCTGGAGGAGATGAGCTCGCGCGCAAGTGGAGAGCC 507  
 Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100  
 QY 508 AAGATGATCGCGCGCATCGCGGCTTTCATCAAGGTGGCGCGCTAGTACGACAGATCTCGATC 567  
 Db 101 LysMetIleGlyGlyIleGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120  
 QY 568 GAGATCTCGCGCAAGNAGGCCATCGGCACCGTGTGATCGCGCGCCACCGCTGAACATC 627  
 Db 121 AspIleCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140  
 QY 628 ATCGCGCGCAACATGCTGACCGCATCGCGCTGCACCTTGAACCTTCCCATCAGCCCATC 687  
 Db 141 IleGlyArgAsnLeuLeuThrGluIleGlyCysThrLeuAsnPheProIleSerProIle 160  
 QY 688 GAGACCGTGCCTGGAAGCTGAAGCCCGCGCATGAGCGCGCCCAAGGTGAAGCAGTGGCCC 747  
 Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180  
 QY 748 CTGACCGGAGAGAGATCAAGCGCTGACCGCATCTCGGAGGAGATCGAGAGAGAGGCG 807  
 Db 181 LeuThrGluGluLysIleLysAlaLeuValGluIleCysThrGluMetGluLysGluGly 200  
 QY 808 AAGATCACCAAGATCGCGCGCGAGAACCCCTTACAAACACCCCGCTGTTCGCCATCAAGAAG 867  
 Db 201 LysIleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 220  
 QY 868 AAGGACGACCAAGTGGCGCAAGCTGTGACTTCGCGAGCTGAAACAAGCGCACCCAG 927  
 Db 221 LysAspSerThrLysTrpArgLysLeuValAspPheArgGluLeuAsnArgLysThrGln 240  
 QY 928 GACTTCTGGAGGTGCAGCTGGGCATCCCGACCGCGCGCGCTGAAGAAGAAGAGAGC 987  
 Db 241 AspPheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSer 260  
 QY 988 GTGACCGTGTGGAGCTGGCGAGCGCTTACTTACAGCTGCCCGCTGGAGAGAGACTTCGCG 1047  
 Db 261 ValThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspLysAspPheArg 280  
 QY 1048 AAGTACACCGCTTACCATCCCGAGCATCAACACGAGACCCCGCGCATCCGCTACCAG 1107  
 Db 281 LysTyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgTyrGln 300  
 QY 1108 TACAACGTGCTGCCCGCGAGGTGAGAGCGCGAGCGCGCGCATCTTCCAGAGAGCATGACC 1167  
 Db 301 TyrAsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThr 320  
 QY 1168 AAGATCTCTGGAGCGCTTCCCGCGCGCAACCCCGAGATCGTATCTACAG-----GCC 1221  
 Db 321 LysThrLeuGluProPheArgLysGlnAsnProAspIleIleIleTyrGlnTyrMetAsp 340  
 QY 1222 CCCCTGTACGTGGCGAGCGACCTCGAGATCGCGCAGACCGCGCGCAAGATCGAGAGCTG 1281  
 Db 341 AspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeu 360  
 QY 1282 CGCAAGACCTCTGCGCTGGGGTTCACACCGCTTCCACCGCGCGCAAGAGACCAAGAGAGCC 1341  
 Db 361 ArgGlnHisLeuLeuLysTrpGlyPheThrThrProAspLysLysHisGlnLysGluPro 380  
 QY 1342 CCCTTCCTGT-----CCCATCGAGCTGCACCGCGCAAGTGGAGCGCTGCAGCCCATCGAG 1395

```
Db 381 ProPheLeuTrpMetGlyTyrGluLeuHisProAspLysTrpThrValGlnProIleVal 400
Qy 1396 CTGCCCGAAGAGGAGCTGGACCGTGAACGACATCCAGAAAGCTGGTGGGCAAGCTGAAC 1455
Db 401 LeuProGlnLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsn 420
Qy 1456 TGGGCCAGCAGATCTACCCCGGATCAAGTGGCGCAGCTGTGCAAGCTCTCTGGCGGC 1515
Db 421 TrpAlaSerGlnIleTyrAlaGlyIleLysValLysGlnLeuCysLysLeuLeuArgGly 440
Qy 1516 GCCAAGGCCCTGACCGCATCGTCCCTGACCGAGGAGGCGGAGCTGGAGCTGGCCGAG 1575
Db 441 ThrLysAlaLeuThrGluValIleProLeuThrLysGluAlaGluLeuGluLeuAlaGlu 460
Qy 1576 AACCGCGAGATCTCTGCGAGCCCGTGCACGCGCTGTACTACGACCCCGAGCAAGGACCTG 1635
Db 461 AsnArgGlnIleLeuLysGluProValHisGlyValTyrAspProSerLysAspLeu 480
Qy 1636 GTGCCCGAGATCCAGAGCAGGCGCACGACAGTGGACCTACCGATCTACGAGAGCC 1695
Db 481 IleValGlnIleGlnLysGlnGlyGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 500
Qy 1696 TTCAAGAACCTGAAGACCGGAGTACGCCAAGATCGCACCGCCACCAACGACGCTG 1755
Db 501 PheLysAsnLeuLysThrGlyLysTyrAlaLysThrArgGlyAlaHisThrAsnAspVal 520
Qy 1756 AAGCAGCTGACCGAGGCGCTGCAGAAATCCCATGGAGACATCGTATCTGGGGCAAG 1815
Db 521 LysGlnLeuThrGluAlaValGlnLysIleAlaAsnGluSerIleValIleTrpGlyLys 540
Qy 1816 ACCCCCAAGTTCGCTGCTGCCATCCAGAGGAGACCTGGGAGACTGGTGACCGACTAC 1875
Db 541 IleProLysPheLysLeuProIleGlnLysGluThrTrpGluThrTrpTrpThrGluTyr 560
Qy 1876 TGGCAGGCCACTGGATCCCGAGTGGAGTTCGTGAACCCCGCTGGTGAAGCTG 1935
Db 561 TrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeu 580
Qy 1936 TGGTACACAGCTGGAGAGGAGCCCATCATCGCGCCGAGACTTCTAGCTGGAGCGGCGC 1995
Db 581 TrpTyrGlnLeuGluLysGluProIleValGlyAlaGluThrPheTyrValAspGlyAla 600
Qy 1996 GCCAACCGCAGACCAAGATCGGCAAGCCCGCTACGTGACCGACCGAGCCGCGCAGAG 2055
Db 601 AlaAsnArgGluThrLysLeuGlyLysAlaGlyTyrValThrSerArgGlyArgGlnLys 620
Qy 2056 ATCGTGACCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCCATCCAGCTGGCC 2115
Db 621 ValValSerLeuThrAspThrThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAla 640
Qy 2116 CTGCAGGACAGCGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATC 2175
Db 641 LeuGlnAspSerGlyLeuGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIle 660
Qy 2176 ATCCAGGCCCGCAGCAGAGCAGAGCGAGCTGGTGAACCCAGATCATCCAGCAGCTG 2235
Db 661 IleGlnAlaGlnProAspLysSerGluSerGluLeuValSerGlnIleIleGluGlnLeu 680
Qy 2236 ATCAAGAGGAGAGGTGTACTGAGCTGGTGGTGGCGCCGACAGGGGATCCGGCGGCAAC 2295
Db 681 IleLysLysGluLysValTyrLeuAlaTrpValProAlaHisLysGlyIleGlyAsn 700
Qy 2296 GAGCAGATCGACAAAGCTGTGAGCAAGGGCATCCGCAAGGTGCTTCTCGAGCGGCATC 2355
Db 701 GluGlnValAspLysLeuValSerAlaGlyIleArgValLeuPheLeuAspGlyIle 720
Qy 2356 GAT 2358
Db 721 Glu 721
RESULT 10
US-10-325-468-20
```

```
; Sequence 20, Application US/10325468
; Publication No. US20040101823A1
; GENERAL INFORMATION:
; APPLICANT: Soong, Nay Wei
; APPLICANT: Pekrun, Katja
; APPLICANT: Shibata, Riri
; TITLE OF INVENTION: HIV-1 VIRAL VARIANTS FOR IMPROVED ANIMAL
; TITLE OF INVENTION: MODELS OF HIV-1 PATHOGENESIS
; FILE REFERENCE: 0166.210US
; CURRENT APPLICATION NUMBER: US/10/325,468
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US 60/343,524
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant / chimeric sequence: clone 1.27 protein Pol
US-10-325-468-20
```

```
Alignment Scores:
Pred. No.: 1,03e-156 Length: 1003
Score: 3384.00 Matches: 634
Percent Similarity: 94.04% Conservative: 44
Best Local Similarity: 87.93% Mismatches: 35
Query Match: 74.29% Indels: 8
DB: 16 Gaps: 4
```

US-09-610-313B-32 (1-2457) x US-10-325-468-20 (1-1003)

```
Qy 220 TTCTTCGCGAGGACCTGGCTTCCCGAGGCAAGCCCGCGAGTTCGCCAGCGAGCAG 279
Db 1 PhePheArgGluAspLeuAlaPheProGlnGlyLysAlaArgLysPheSerSerGln 20
Qy 280 AACCGCGCAACAGCCCCACGCGAGCTGCAGGTG-----CGCGGCGCAACACCC 333
Db 21 ThrArgAlaAsnSerProIleArgGluArgGlnValTrpArgArgAspAsnAsnSer 40
Qy 334 CGCAGCGAGCGCGCGCGAGCGCGAGCGCGAGCGCGAGCTG-----AACTTCCCGAGTACC 387
Db 41 LeuSerGluAlaGlyAlaAspArgGlnGlyThrValSerPheAsnPheProGlnIleThr 60
Qy 388 CTGTGGCAGCGCCCGCTGGTGAGCATCAAGTGGCGCGCGAGTCAAGAGAGGCGCTGCTG 447
Db 61 LeuTrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeu 80
Qy 448 GACACCGCGCGCGAGCACCGCTGCTCGAGAGATGAGCTGCGCGCAAGTGAAGGCC 507
Db 81 AspThrGlyAlaAspAspThrValLeuGluAspMetAspLeuProGlyArgTrpLysPro 100
Qy 508 AAGATGATCGCGCGCTCGCGGCTTTCATCAGGTGGCGCGAGTACGACCATCTGATC 567
Db 101 LysMetIleGlyGlyIleGlyGlyPheIleLysValArgGlnTyrAspGlnIleProIle 120
Qy 568 GAGATCTGCGCAAGAAGGCCATCGCACCGCTGCTGATCGCGCCCGCGCGCGTGAACATC 627
Db 121 AspileCysGlyHisLysAlaValGlyThrValLeuValGlyProThrProValAsnIle 140
Qy 628 ATCGCGCGCAACATGTGACCCAGCTGGGTGACCCCTGACCTTCCCATCAGCCCCATC 687
Db 141 IleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIle 160
Qy 688 GAGACCGTGGCGTGAAGCTGAAGCCCGCGCATGACCGCGCCCAAGGTGAAGCAGTGGCCC 747
Db 161 GluThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnTrpPro 180
Qy 748 CTGACCGGAGGAGATCAAGGCCCTGACCGCCATCTGCGAGGAGATGAGAGAGGAGGC 807
Db 181 LeuThrGluGluLysIleLysAlaLeuValIleCysThrGluMetGluLysGluGly 200
```



Db 21 ThrGlyAlaAenSerSerAlaSerArgLysLeu-----GlyAspGly----- 34  
Qy 340 GAGCGCGCGCGAGCGCGAGCGCACC-----CTGAATCTTCCCGCCAGATCACCCCTG 390  
Db 35 ---GlyGlyAlaGluArgGlnGlyThrSerSerSerPheSerPheProGlnIleThrLeu 53  
Qy 391 TGGCAGCGCCCTCTGGTGGAGCATCAAGGTGGCGGCGCCAGATCAAGAGGCGCTCTCGAC 450  
Db 54 TrpGlnArgProLeuValThrIleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp 73  
Qy 451 ACCGCGCGCAGCAGCAGCGCTGGAGGAGATGAGCTGCCCGGCAAGTGAAGCCCAAG 510  
Db 74 ThrGlyAlaAspThrValLeuGluAspIleAsnLeuProGlyLysIleThrLysProLys 93  
Qy 511 ATGATCGCGCGCATCGCGGCTTCAATCAAGGTGGCGCAGTAGCAGCAGATCCTCATCGAG 570  
Db 94 MetIleGlyIleGlyGlyPheIleLysValArgGlnThrAspGlnIleLeuIleGlu 113  
Qy 571 ATCTGGCGCAAGAGCGCATCGCACCGTGTGATCGCGCCCGCCAGTGAACATCATC 630  
Db 114 IleCysGlyLysLysAlaIleGlyThrValLeuValGlyProThrProValAsnIleIle 133  
Qy 631 GGCGCGACATCTGACCCAGCTGGCTGGCTGACCTGACCTGAATCTCCCATCAGCCCATCGAG 690  
Db 134 GlyArgAsnMetLeuThrGlnIleGlyCysThrLeuAsnPheProIleSerProIleAsp 153  
Qy 691 ACCGTGCGGTGAAGTGAAGCGCGATGACCGCGCCCAAGGTGAAGCATGTGCCCTG 750  
Db 154 ThrValProValLysLeuLysProGlyMetAspGlyProLysValLysGlnIleProLeu 173  
Qy 751 ACCGAGGAGAGATCAAGCGCTGACCGCCATCTCGCGAGGAGATGAGAGAGGCGCAAG 810  
Db 174 ThrGluGluLysIleLysAlaLeuThrGluIleCysLysGluMetGluGluGluLys 193  
Qy 811 ATCACCAGATCGCGCGCGAGAACCCCTACAACACCCCGCTGTTCGGCATCAAGAGAAG 870  
Db 194 IleSerLysIleGlyProGluAsnProTyrAsnThrProValPheAlaIleLysLys 213  
Qy 871 GACAGCACCAAGTGGCGCAAGCTGGTGAAGTTCGCGAGCTGAACAGAGCGCACCAGGAC 930  
Db 214 AspSerThrLysIleLysLeuValAspPheArgGluLeuAsnLysArgThrGlnAsp 233  
Qy 931 TTCTGGAGGTGACGTGGCGCATCCCCACCGCGCGCTGAAGAGAGAGAGAGCGGTG 990  
Db 234 PheTrpGluValGlnLeuGlyIleProHisProAlaGlyLeuLysLysLysSerVal 253  
Qy 991 ACCGTGTGACGTGGCGGACGCTTACTTACGTGCGCTGGAGCGAGCATCTCCGCAAG 1050  
Db 254 ThrValLeuAspValGlyAspAlaTyrPheSerValProLeuAspGluSerPheArgLys 273  
Qy 1051 TACACCGCTTACCATCCCGCATCAACAGCAGACCCCGCGCATCCGCTACCGATAC 1110  
Db 274 TyrThrAlaPheThrIleProSerIleAsnAsnGluThrProGlyIleArgThrGlnTyr 293  
Qy 1111 AACGTGTGCGCCAGGCGTGAAGGCGAGCGCCAGCATCTTCACAGCAGCATCACCAAG 1170  
Db 294 AsnValLeuProGlnGlyTrpLysGlySerProAlaIlePheGlnSerSerMetThrLys 313  
Qy 1171 ATCTGTAGGCGCTTCCCGCGCGCGCAACCGCGAGATCGTATCTACCAAG-----GCCCGC 1224  
Db 314 IleLeuGluProPheArgIleLysAsnProGluMetValIleTyrGlnTyrMetAspAsp 333  
Qy 1225 CTGTACCTGGCGCAGCATCGGATCGCGCAGCAGCGCGCCAGATCGAGGAGCTCGCG 1284  
Db 334 LeuTyrValGlySerAspLeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArg 353  
Qy 1285 AAGCACCTGTCTGGCGGCTTCAACACCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1344  
Db 354 AlaHisLeuLeuArgTrpGlyPheThrThrProAspLysLysHisGlnLysGluProPro 373  
Qy 1345 TTCTCTG-----CCCATCGAGCTGCACCCCGCAGCAAGTGGACCGGTGCAGCCATCGAGCTG 1398  
|||||

Db 374 PheLeuTrpMetGlyTyrGluLeuLeuHisProAspArgTrpThrValGlnProIleGluLeu 393  
Qy 1399 CCCGAGAGAGAGAGCTGCAGCCGTCAACAGCATCAGAGCTGGTGGGCAAGCTGAACCTGG 1458  
Db 394 ProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrp 413  
Qy 1459 GCCAGCCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGCAAGCTGCTCGCGGCGCC 1518  
Db 414 AlaSerGlnIleTyrAlaGlyIleLysValLysGlnLysLeuCysLysLeuLeuArgGlyThr 433  
Qy 1519 AAGCCCTGACCGACATCGTGGCCCTGACCGAGGAGCGCGAGCTGGAGCTGGCGGAGAAC 1578  
Db 434 LysAlaLeuThrAspIleValProLeuThrGluGluAlaGluLeuGluGluAsn 453  
Qy 1579 CGCGAGATCTTCGCGAGCGCGCTGACGGCTGCTACTACGACCCCGCAGCAAGCAAGCTGTG 1638  
Db 454 ArgGluLeuLeuArgGluProValHisGlyValTyrTyrAspProSerLysAspLeuVal 473  
Qy 1639 GCCGAGATCCAGAGCAGGCGCACGACGATGGACCTTACCGAGATCTTACGAGCCCTTC 1698  
Db 474 AlaGluValGlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyrGlnGluProPhe 493  
Qy 1699 AAGAACCTGAGAGCCGCGCAAGTAGTACCCCAAGATGCGCACCGCCCGCAGCAAGCAAGCTG 1758  
Db 494 LysAsnLeuLysThrGlyLysTyrSerArgLysArgSerAlaHisThrAsnAspValArg 513  
Qy 1759 CAGCTGACCGAGCGCTGCAGAGATCGCATGAGAGCATCGTATCTGGGGCAAGACC 1818  
Db 514 GlnLeuThrGluValValGlnLysIleAlaThrGluSerIleValIleTrpGlyLysThr 533  
Qy 1819 CCAGAGTTCGCGCTGCCATCCAGAGGAGACCTGGGAGACCTGGTGGACCGACTACTGG 1878  
Db 534 ProlLysPheArgLeuProLysGlnArgGluThrTrpGluThrTrpTrpMetGluTyrTrp 553  
Qy 1879 CAGCGACCTGGATCCCGAGTGGAGTTGCGTGAACACCCCGCTGGTGAAGCTGTGG 1938  
Db 554 GlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeuValLysLeuTrp 573  
Qy 1939 TACCAGCTGAGAGAGGAGCCCATCATCGCGCGCGAGACCTTCTACGTGGAGCGCGCGCC 1998  
Db 574 TyrGlnLeuGluLysAspProlIleValGlyAlaGluThrPheTyrValAspGlyAlaAla 593  
Qy 1999 AACCGCGAGACCAAGATCGGAGCGCGGTACGTGACCGACCGCGCGCGCGCAGAGATC 2058  
Db 594 SerArgGluThrLysLeuGlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysVal 613  
Qy 2059 GTGAGCTGACCGAGACCAACAGAGACCGAGCTGAGCGCATCCAGCTGCGCGCTG 2118  
Db 614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
Qy 2119 CAGGACAGCGCGCAGGAGTGAACATCGTACCGACAGCAGCTACGCTGGGCGCATCATC 2178  
Db 634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653  
Qy 2179 CAGCGCCAGCCGCAAGAGCAGGAGCGAGCTGTGTAACCCAGATCATCGAGCAGCTGATC 2238  
Db 654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleGluGluLeuIle 673  
Qy 2239 AAGAAGAGAGAGTGTACTGAGCTGGGTGGTCCCGCCCGCAGAGGCGATCGCGCGCAAG 2298  
Db 674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyLysAsnGlu 693  
Qy 2299 CAGATCGACAGCTGGTGAAGCAAGGCGATCCGCAAGGTGTCTTCTGGAGCGGATCGAT 2358  
Db 694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713  
RESULT 12  
US-10-296-734-1470  
; Sequence 1470, Application US/10296734  
; Publication No. US20040054137A1  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Scott A  
; APPLICANT: Ramshaw, Ian A

; TITLE OF INVENTION: Synthetic molecules and uses therefor

; FILE REFERENCE: Savine

: CURRENT APPLICATION NUMBER: US/10/296.734

CURRENT APPLICATION NUMBER: 09/1  
: CURRENT FILING DATE: 2003-08-04

; CURRENT FILING DATE: 2003-08-04  
 : PRIOR APPLICATION NUMBER: AII P07761/A00

;; PRIOR APPLICATION NUMBER: 3000 05

; PRIOR FILING DATE: 2000-05-

; NUMBER OF SEQ ID NOS: 1507

; SOFTWARE: Pat

SEQ	ID	NO	14
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28
29	30	31	32
33	34	35	36
37	38	39	40
41	42	43	44
45	46	47	48
49	50	51	52
53	54	55	56
57	58	59	60
61	62	63	64
65	66	67	68
69	70	71	72
73	74	75	76
77	78	79	80
81	82	83	84
85	86	87	88
89	90	91	92
93	94	95	96
97	98	99	100

; LENGTH: 95

/ LENGTH: 333  
: TYPE: PRT; TYPE: PRI  
; ORGANISM: Human

**Alignment Scores:**

Alignment Scores:				
Pred. No.:	1,21e-155	Length:	995	
Score:	632.00	Matches:	634	
Percent Similarity:	93.61%	Conservative:	40	
Best Local Similarity:	88.06%	Mismatches:	32	
Query Match:	73.81%	Indels:	14	
DB:	15	Gaps:	5	

US-09-610-313B-32 (1-2457) x US-10-296-734-1470 (1-995)

220	TTC	TCCGCGAGGA	CCTGGCCTTCCCCAGGGCAAGGCCCGGAGTTCCCACGAGCAAG	279
1	Phe	PheArgLysLeuAlaPhe	GlnGlnGlyLysAlaargGluPheSerSerGluGln	20
280	AACC	GCGCCAACAG	CCCCACCAGCGCGGAGCTGCAGGTGCGCGCGGCACAACCCCCCGCACG	339
21	Thr	GlyAlaAsnSerSerAla	SerArgLysLeu-----GlyAspGly-----	34
340	GAGG	CCGCGCGCAGCGCCAGGGCAAC-----CTGAATCTCCCAAGATCACCCCTG	390	
35	--	GlyGlyAlaGluArgGlnGlyThr	SerSerSerPheSerPheProGlnIleThrLeu	53
391	TGG	CAGCGCCCCCTCGTGAGCATCAAG	TGGCGGCCAGATCAAGAGGAGCCCTGCTGGAC	450
54	Trp	GinArgProLeuValThr	IleLysIleGlyGlnLeuLysGluAlaLeuLeuAsp	73
451	ACGG	CGCGCAGCAGCACCGTCTCGGAGGATGACCTG	CGCGCAAGTGGAGCCCAAG	510
74	Thr	GlyAlaAspAspThrVal	LeuGluAspIleAsnLeuProGlyLysTrpLysProLys	93
511	ATG	ATCGCGGGCATCGCGGGCTTCATCAAGTGGCGCAGTAC	GACCAGATCTCTGATCGAG	570
94	Met	IleGlyGlyIleGlyGlyPheIleLysValArgGlnTrp	aspGlnIleLeuIleGlu	113
571	ATT	CTGGCGAAGAAGGCCATCGGCAACGTCGTGATCGG	CCCCACCACCCCGTGAACTATCT	630
114	Ile	CysGlyLysLysAlaIleGlyThrValLeuValGlyPro	ThrProValAsnIleIle	133
631	GSG	CGGACATCTCTGACCCAGCTGGGCTGCACCTGAACT	TCCCATCAGCCCATCGAG	690
134	Gly	ArgAsnMetLeuThrGln	IleGlyCysThrLeuAsnPheProIleSerProIleAsp	153
691	ACGT	GTCGCGTCAAGCTGAAGCCCGGCATGACGCGCC	CAAGTGAAGCAGTGGCGCCCTG	750
154	Thr	ValProValLysLeuLysProGlyMetAspGlyPro	LysLysGlnTrpProLeu	173
751	ACCG	AGGAGAAGATCAAGGCCCTCACCGCCCATCTGCG	GAGGATGGAGAGGAGGCAAG	810
174	Thr	GluGluLysIleLysAlaLeuThrGluIleCysLysGlu	MetGluGluGluLys	193
811	ATC	ACAAGATCGGCGCCCGAGAACCCCTTACAACAC	CCCCCGTGTTCGCCATCAAGAAGAAG	870
194	Ile	SerLysIleGlyProGluAsnProTyrAsnThrPro	ValPheAlaIleLysLysLys	213
871	GAC	AGCACCAAGTGGCGCAACTGGTGACTTCGCGAGCT	GTACACAGCGCACCCAGGAC	930
214	Asp	SerThrLysTrpArgLysLeuValAsp	PheArgGluLeuAsnLysArgThrGlnAsp	233
931	TTT	TGGAGGTGCAGCTGGGCATCCCCCACCCCGCGCCT	CTGAAGAGAGAAAGAGACGCTG	990

594 SerArgGluThrLysLeuGlyLySAlaGlyTyrValThrAspArgGlyAra-gInLysVal 613

2059 GTAGCCTGTACCCGAGACCAACACAGAACCGAGCTGCAGGCCATCCAGCTGGCCCTG 2118  
:::|||||  
614 IleSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIleHisLeuAlaLeu 633  
|||||  
2119 CAGACAGCCGCGACGAGTGTAACATCGTGACCAGCACGACCATGAGCCCTGGGGCATCATC 2178  
|||||  
634 GlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIle 653  
|||||  
2179 CAGSCCCAGCCCGACAAGACGAGCGAGCGAGCTGTGTGAACACGATCATCTGAGCAGCTGATC 2238  
|||||  
654 GlnAlaGlnProAspArgSerGluSerGluValValSerGlnIleIleIleGluLeuIle 673  
|||||  
2239 AAGAAGGAGAGAGGTGTACTCTGAGCTGGTGTCCCCGCCACAAAGGCGATCGCGGCAACGAG 2298  
|||||  
674 LysLysGluLysValTyrLeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGlu 693  
|||||  
2299 CAGATCGACAAGCTGTGTGAGCAAGGGCATCGCAAGGTGCTGTTCCTGGAGCGGATCGAT 2358  
|||:::|||||  
694 GlnValAspLysLeuValIleSerGlyIleArgLysValLeuPheLeuAspGlyIleAsn 713  
|||||

RESULT 13

US-10-296-734-1471

; Sequence 1471, Application US/10296734

; Publication No. US20040054137A1

; GENERAL INFORMATION:

; APPLICANT: Thompson, Scott A

; TITLE OF INVENTION: Synthetic molecules and uses therefor

; FILE REFERENCE: Savine

; CURRENT APPLICATION NUMBER: US/10/296,734

; PRIOR FILING DATE: 2003-08-04

; PRIOR APPLICATION NUMBER: AU PQ7761/00

; PRIOR FILING DATE: 2000-05-26

; NUMBER OF SEQ ID NOS: 1507

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1471

; LENGTH: 1006

; TYPE: PRT

; ORGANISM: Artificial

; FEATURE:

; OTHER INFORMATION: POL Consensus A

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (15)..(15)

; OTHER INFORMATION: X = unknown amino acid

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (22)..(23)

; OTHER INFORMATION: X = unknown amino acid

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (26)..(26)

; OTHER INFORMATION: X = unknown amino acid

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (30)..(30)

; OTHER INFORMATION: X = unknown amino acid

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (36)..(36)

; OTHER INFORMATION: X = unknown amino acid

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (38)..(39)

; OTHER INFORMATION: X = unknown amino acid

; FEATURE:

; NAME/KEY: MISC FEATURE

; LOCATION: (41)..(44)

; OTHER INFORMATION: X = unknown amino acid

; FEATURE:

; NAME/KEY: MISC FEATURE





Db 281 \*\*\*PheArgLysTyrThrAlaPheThrIleProSer\*\*\*AsnAsnGluThrProGly\*\*\* 300  
 QY 1099 CGCTACAGTACACGCTGCTGCCAGCGGTGGAAGGGCCAGCCCGACCATCTTCCAGAGC 1158  
 Db 301 ArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerPro\*\*\*IlePheGln\*\*\* 320  
 QY 1159 AGCATGACCAAGATCTCGAGCCCTTCCGGCCCGCAACCCCGAGATCGTGATCTACAG 1218  
 Db 321 SerMetThrLysIleLeuGluProPheArg\*\*\*\*\*Pro\*\*\*IleValIleTyrGln 340  
 QY 1219 -----GCCCCCTGACGTGGGCGAGCTGGAGATCGCCAGCACCGCGCCAGATC 1272  
 Db 341 TyrMetAspAspLeuTyrValGlySerAspLeuGluIleGlyGlnHisArgAlaLysIle 360  
 QY 1273 GAGGAGCTGGCGAGCACCTGCTCGCTGGCGCTTCCACCCCGGACCAAGAGACCCAG 1332  
 Db 361 GluGluLeuArg\*\*\*HisLeuLeu\*\*\*TrpGlyPhe\*\*\*ThrProAspLysLysHisGln 380  
 QY 1333 AAGAGCCCTCTCTCTG-----CCCATCGAGCTGCACCCCGACCAAGTGGACCGTGCAG 1386  
 Db 381 LysGluProProPheLeuTyrMetGlyTyrGluLeuHisProAspLysTyrThrValGln 400  
 QY 1387 CCCATCGAGCTGCCGAGAGGAGCTGCACCGTGAACGATCCAGAGCTGGTGGC 1446  
 Db 401 Pro\*\*\*\*\*LeuProGluLysAspSerTrpThrValAsnAspIleGlnLysLeuValGly 420  
 QY 1447 AAGCTGAATCTGGGCGACCATCTACCCCGCATCAAGGTGCCCGACCTGTGCAAGCTG 1506  
 Db 421 LysLeuAsnTrpAlaSerGlnIleTyrAlaGlyIleLys\*\*\*LysGlnLeuCys\*\*\*Leu 440  
 QY 1507 CTGGCGGCGCCAAAGCCCTGACGACATCTGCTGCCCTGACGAGGAGCGCGAGCTGAG 1566  
 Db 441 LeuArgGlyAlaLysAlaLeuThrAspIleVal\*\*\*LeuThrGluGluAlaGluLeuGlu 460  
 QY 1567 CTGGCGGAGAACCGGAGATCTCGCGAGCGCGTGCACCGCTGTACTAGACCCCGCAGC 1626  
 Db 461 LeuAlaGluAsnArgGluIleLeuLys\*\*\*ProValHisGlyValTyrAspPro\*\*\* 480  
 QY 1627 AAGCACTGTGTGGCGGAGATCTCAGAACGAGCGGCGCACAGTGGACCTACCATGATCTAC 1686  
 Db 481 LysAspLeuValAlaGlu\*\*\*GlnLysGlnGlyGlnAspGlnTrpThrTyrGlnIleTyr 500  
 QY 1687 CAGGAGCCCTCAAGAACCTGAAGACCGCAAGTACGCCAAGATGCGCACCGCCACACC 1746  
 Db 501 GlnGluProPheLysAsnLeuLysThrGlyLysTyrAla\*\*\*LysArgSerAlaHisThr 520  
 QY 1747 AAGCACTGAAGCACTGACCGAGCGCGTGCAGAGATCCCATGGAGCATCTGATC 1806  
 Db 521 AsnAspValLysGlnLeuThrGluValValGlnLysVal\*\*\*\*\*GluSerIleValIle 540  
 QY 1807 TGGGGCAAGACCCCAAGTTCCGCTCCCATCCAGAAAGGAGACCTGGGAGACCTGTGG 1866  
 Db 541 TrpGlyLys\*\*\*ProLysPheArgLeuProIleGln\*\*\*GluThrTrpGlu\*\*\*TrpTrp 560  
 QY 1867 ACCGACTACTGGCAGGCCACCTCGATCCCGAGTGGGAGTTCGTGAACACCCCGCCCTG 1926  
 Db 561 MetGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheValAsnThrProProLeu 580  
 QY 1927 GTGAAGCTGTGTACAGCTGGAGAGAGCCCATCATCGCGCGGAGACCTTCTAGCTG 1986  
 Db 581 ValLysLeuTrpTyrGlnLeuGluLysAspProIle\*\*\*GlyAlaGluThrPheTyrVal 600  
 QY 1987 GACGGCGCCCAACCGGAGACCAAGATCGCAAGCGCGCTACGTGACCGACCGCGGGC 2046  
 Db 601 AspGlyAlaAlaAsnArgGluThrLys\*\*\*GlyLysAlaGlyTyrValThrAspArgGly 620  
 QY 2047 CGGCAGAGATCTGTGACCTGACCGAGACCAACCAAGAGACCGAGCTGCAGGCGCATC 2106  
 Db 621 ArgGlnLysValValSerLeuThrGluThrThrAsnGlnLysThrGluLeuHisAlaIle 640  
 QY 2107 CAGCTGCGCTCGAGGACGCGGAGGAGGTGAACATCGTGAACCGGACCGCATGACGCC 2166

Db 641 HisLeuAlaLeuGlnAspSerGlySerGluValAsnIleValThrAspSerGlnTyrAla 660  
 QY 2167 CTGGGATCATTCAGGCCCCAGCCAGACAGAGCGAGAGCTGTGTGAACACGATCATC 2226  
 Db 661 LeuGlyIleIleGlnAlaGlnProAspArgSerGluSerGlu\*\*\*ValAsnGlnIleIle 680  
 QY 2227 GAGCAGCTGATCAAGAGAGGAGAGGTGTACTGAGCTGGGTGCCCGCCCAAGGGGATC 2286  
 Db 681 GluLysLeuIle\*\*\*Lys\*\*\*LysValTyrLeuSerTrpValProAlaHisLysGlyIle 700  
 QY 2287 GGCGGCAACAGCAGCAGATCGAACCTGTGAGCAAGGGCATCCGCAAGGTGCTCTCTG 2346  
 Db 701 GlyGlyAsnGluGlnValAspLysLeuValSer\*\*\*GlyIleArgLysValLeuPheLeu 720  
 QY 2347 GACGCGATCGAT 2358  
 Db 721 AspGlyIleAsp 724

RESULT 14  
 US-10-301-661A-6  
 ; Sequence 6, Application US/10301661A  
 ; Publication No. US20030157660A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE-  
 ; APPLICANT: INSERM  
 ; APPLICANT: ASSISTANCE PUBLIQUE-HOPITAUX DE PARIS  
 ; APPLICANT: INSTITUT PASTEUR  
 ; APPLICANT: MAUCLERE, Philippe  
 ; APPLICANT: LOUSSERT-AJAKA, Ibtissam  
 ; APPLICANT: SIMON, Francois  
 ; APPLICANT: SARAGOSTI, Sentob  
 ; APPLICANT: BARRE-SINOSSI, Francoise  
 ; TITLE OF INVENTION: NON-M NON-O HIV STRAINS, FRAGMENTS AND APPLICATIONS.  
 ; FILE REFERENCE: 598US12  
 ; CURRENT APPLICATION NUMBER: US/10/301,661A  
 ; CURRENT FILING DATE: 2002-11-22  
 ; PRIOR APPLICATION NUMBER: US/09/319,588C  
 ; PRIOR FILING DATE: 1999-08-27  
 ; PRIOR APPLICATION NUMBER: FR96/15087  
 ; PRIOR FILING DATE: 1996-12-09  
 ; NUMBER OF SEQ ID NOS: 98  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 1014  
 ; TYPE: PRT  
 ; ORGANISM: Human immunodeficiency virus type 1  
 US-10-301-661A-6

Alignment Scores:  
 Pred. No.: 7,22e-145 Length: 1014  
 Score: 3140.50 Matches: 579  
 Percent Similarity: 89.70% Conservative: 74  
 Best Local Similarity: 79.53% Mismatches: 60  
 Query Match: 68.95% Indels: 15  
 DB: 14 Gaps: 5

US-09-610-313B-32 (1-2457) x US-10-301-661A-6 (1-1014)

QY 220 TTCTTCGGGAGACCTGGCTTCCCGGCAAGCGCCGAGTTCGCCAGCGAGCAG 279  
 Db 1 PhePheArgGluGluGlnValSerLeuGluArgGluThrArgLysLeuProAspAsn 20  
 QY 280 AAC-----CGCCCAACACGCCAGCCAGCGCGAGCTGCAGGTG----- 318  
 Db 21 AsnLysGluArgAlaHisSerProAlaThrArgGluLeuTrpValSerGlyGlyGluGlu 40  
 QY 319 ---CGCGGCGCAACCCCGCAGCGCGCGCGCGAGCGCCAG-----GGCACC 366  
 Db 41 HisThrGlyGluGlyAspAlaGlyGluProGlyGluAspArgGluLeuSerValProThr 60  
 QY 367 CTGAATCTCCCGCAGATCATCCCTGTGCGAGCGCCCTGTGAGCATCAAGTGGCGGC 426  
 Db 61 PheAsnPheProGlnIleThrLeuTrpGlnArgProValIleThrValLysIleGlyLys 80

QY 427 CAGATCAAGGAGCCCTGCTGACACCGCGCCGACGACACCGTCTGGAGGAGATGAC 486  
 Db 81 GluValArgGluAlaLeuLeuAspThrGlyAlaAspAspThrValIleGluLeuGln 100  
 QY 487 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCAATCAAGTGGCG 546  
 Db 101 LeuGluGlyLeuTrpLysProLysMetIleGlyGlyIleGlyPheIleLysValArg 120  
 QY 547 CAGTACGACCATCTGATCAGATCTGCGGCAAGAGAGCGCATCGGACCGTGTGATC 606  
 Db 121 GlnTyrAspAsnIleThrValAspIleGlnGlyArgLysAlaValGlyThrValLeuVal 140  
 QY 607 GCGCCCAACCCCGTGAACATCATCGCGCGCAACATCTGACCCACCTGGCTGGCCCTG 666  
 Db 141 GlyProThrProValAsnIleIleGlyArgAsnLeuLeuThrGlnIleGlyCysThrLeu 160  
 QY 667 AACTTCCCATCAGCCCATCAGACCGTGGCCGCTGAAGCTGAAGCCCGGATGACGCGC 726  
 Db 161 AsnPheProIleSerProIleGluThrValProValLysLeuLysProGlyMetAspGly 180  
 QY 727 CCCAAGTGAAGCAGTGGCCCTGACGAGGAGAGATCAAGGCCCTGACCGCCATCTGC 786  
 Db 181 ProLysValLysGlnTrpProLeuThrThrGluLysIleGluAlaLeuArgGluIleCys 200  
 QY 787 GAGGAGATGGAGAGGCGGCAAGATCACCAAGATCGCGCCCGGACACCCCTACACACC 846  
 Db 201 ThrGluMetGluLysGluLysIleSerArgIleGlyProGluAsnProTyrAsnThr 220  
 QY 847 CCGCTGTTCCGCATCAAGAAGAGACACACCAAGTGGCGCAAGCTGGTGGACTTCGCG 906  
 Db 221 ProIlePheAlaIleLysLysLysAspSerThrLysTrpArgLysLeuValAspPheArg 240  
 QY 907 GAGCTGAACACCGCACCGACATCTGCGAGAGTGCAGTGGGATCCCCACCCCGCC 966  
 Db 241 GluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeuGlyIleProHisProAla 260  
 QY 967 GCGCTGAAGAGAGAGCGTACCGTCTGGAGCGTGGCGCGCTACTTCTGACGCTG 1026  
 Db 261 GlyLeuLysGlnLysSerValThrValLeuAspValGlyAspAlaTyrPheSerCys 280  
 QY 1027 CCGCTGGACGAGGACTTCGCAAGTACACCGCTTCCACCATCCCCAGATCAACACCGAG 1086  
 Db 281 ProLeuAspLysAspPheArgLysTrpAlaPheThrIleProSerIleAsnAsnGlu 300  
 QY 1087 ACCCGCGGATCCGCTACAGTACAACTGCTGCGCCCGGCGTGAAGGCGACCCCGAC 1146  
 Db 301 ThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGlyTrpLysGlySerProAla 320  
 QY 1147 ATCTTCCAGACGATGACCAAGATCTGGAGCGCTTCCGCGCCGCAACCCCGAGATC 1206  
 Db 321 IlePheGlnSerThrMetThrLysIleLeuGluProPheArgGluLysPheTrpThrProAsp 340  
 QY 1207 GTGATCTACCG- - - - -GCCCGCTGTACGTGGCGACGACTGGAGATCGCCGACGAC 1260  
 Db 341 IleIleTyrGlnTyrMetAspAspLeuTyrValGlySerAspLeuGluLeuAlaGlnHis 360  
 QY 1261 CGCGCCCAAGATCGAGAGTGGCGAAGCATCTGCGTGGCGCTTCCACACCCCGAC 1320  
 Db 361 ArgGluAlaValGluAspLeuArgAspHisLeuLeuLysTrpGlyPheThrThrProAsp 380  
 QY 1321 AAGAGACCAAGAGGCGCCCTTCTG- - - - -CCCATCGAGCTGACCCCGACAG 1374  
 Db 381 LysLysHisGlnLysGluProProPheLeuTrpMetGlyTyrGluLeuHisProAspLys 400  
 QY 1375 TGGACCGTGCAGCCCATCGAGTGGCCGAGAGAGAGTGCACCGTGAACACATCCAG 1434  
 Db 401 TrpThrValGlnProIleLysLeuProGluLysAspValTrpThrValAsnAspIleGln 420  
 QY 1435 AAGCTGGTGGCAAGCTGAACCTGGCGCGCCAGATCTACCCCGGATCAAGGTGGCGAC 1494  
 Db 421 LysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrProGlyIleArgValLysGln 440

QY 1495 CTGTGCAAGTGTGCTGCGCGCGCAAGCCCTGACCGACATCTGCTGCCCTGACCGAGAG 1554  
 Db 441 LeuCysLysLeuIleArgGlyAlaArgAlaLeuThrGluValValAsnPheThrGluGlu 460  
 QY 1555 CCGAGCTGGAGCTGGCCGAGAACCGCGAGATCTTCCGCGAGCCCGTGCACGGCGTGTAC 1614  
 Db 461 AlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGluProLeuHisGlyValTyr 480  
 QY 1615 TACGACCCCAACAGGACCTGTGCGCGAGATCCAGAGACGAGGCGCACGACGAGTGACC 1674  
 Db 481 TyrAspProGlyLysGluLeuValAlaGluIleGlnLysGlnGlyGlnGlyValTrpThr 500  
 QY 1675 TACGAGATCTACAGGAGCCCTTCAAGAACCTGAAGACCGCAAGTACCGCAAGATGCGC 1734  
 Db 501 TyrGlnIleTyrGlnGluLeuHisLysAsnLeuLysThrGlyLysTyrAlaLysMetArg 520  
 QY 1735 ACCGCGCCACCAACGAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCGCCATGGAG 1794  
 Db 521 SerAlaHisThrAsnAspIleLysGlnLeuValGluValValArgLysValAlaThrGlu 540  
 QY 1795 AGCATCTGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGGAGACCTGG 1854  
 Db 541 SerIleValIleTrpGlyLysThrProLysPheArgLeuProValGlnLysGluValTrp 560  
 QY 1855 GAGACCTGGTGGACCGACTACTTGGCAGCGCACCTGGATCCCGAGTGGGAGTTCGTGAAC 1914  
 Db 561 GluAlaTrpTrpThrAspHisTrpGlnAlaThrTrpIleProGluTrpGluPheValAsn 580  
 QY 1915 ACCCGCCCGCTGGTGAAGCTGTGTGTACAGCTGGAGAGAGCCCATCATCGCCCGGAG 1974  
 Db 581 ThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGluProIleSerGlyAlaGlu 600  
 QY 1975 ACCTTCTACGTGGAGCGCGCCCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGTG 2034  
 Db 601 ThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyPheVal 620  
 QY 2035 ACCGACCGCGCGCGCAGAAGATCGTGTGCTGACCGAGACCCAGACCAACAGAGAGCCGAG 2094  
 Db 621 ThrAspArgGlyArgGlnLysValValSerIleAlaAspThrThrAsnGlnLysAlaGlu 640  
 QY 2095 CTGACGCGCATCTGAGCTGGCCCTGCGAGCACGCGCAGCGAGGTGAACATCGTGCACGAC 2154  
 Db 641 LeuGlnAlaIleLeuMetAlaLeuGlnGluSerGlyArgAspValAsnIleValThrAsp 660  
 QY 2155 AGCCAGTACGCGCTGGGATCATCTCAGCGCCAGCGCGCAAGAGCGAGCGAGTGTG 2214  
 Db 661 SerGlnTyrAlaMetGlyIleIleHisSerGlnProAspLysSerGluSerGluLeuVal 680  
 QY 2215 AACGAGATCATCGAGCAGCTCATCAAGAGAGAGAGGTGTACCTGAGTGGTGGCCGCGC 2274  
 Db 681 SerGlnIleIleGluGluLeuLysLysGluArgValTyrLeuSerTrpValProAla 700  
 QY 2275 CACAAGGCGATCGCGCGCAACGAGCAGATCGACAAGCTGTGTGAGCAAGGCGATCGCGAAG 2334  
 Db 701 HisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuValSerSerGlyIleArgLys 720  
 QY 2335 GTGCTGTCTCTGGAGCGCATCATGAT 2358  
 Db 721 IleLeuPheLeuAspGlyIleGlu 728  
 RESULT 15  
 US-09-952-060-35  
 ; Sequence 35, Application US/09952060  
 ; Publication No. US2003004421A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Emini, Emilio A.  
 ; APPLICANT: Youil, Rima  
 ; APPLICANT: Bett, Andrew J.  
 ; APPLICANT: Chen, Ling  
 ; APPLICANT: Kaslow, David C.  
 ; APPLICANT: Shiver, John W.  
 ; APPLICANT: Toner, Timothy J.  
 ; APPLICANT: Casimiro, Danilo R.

QY	13	ATGGCGGAGGCCATGAGCCAGCCACC---	AGCGCCAAACATCTCTGATGACGCGCAGCAAC	69
Db	363	LeuAlaGluAlaMetSerGlnValThrAsnSerAlaThrIleMetMetGlnArgGlyAsn	382	
QY	70	TTCAAGGGCCCCAAGCGGATCATCAAGTGCTTCAACTGCGCAGGAGGAGGCGCCACATCGGC	129	
Db	383	PheArgAsnGlnArgGlyThrValLysCysPheAsnCysGlyLysValGlyLysIleAla	402	
QY	130	CGCAACTGCGCGCCCCCGCAAGAGGGCTGCTGGAAGTCGCGCAAGGAGGGCCACCAAG	189	
Db	403	LysAsnCysArgAlaProArgLysLysGlyCysIrpLysCysGlyLysGluGlyHisGln	422	
QY	190	ATGAGGACTGCAACGAGGGCCAGGCCAACTTCTTCGCGAGAGACTGGCTTCCCCCAG	249	
Db	423	MetLysAspCysAsnGluArgGlnAlaAsnPheLeu-GlyLysIleIrpProSerHisLys	442	
QY	250	GGCAGGGCCCGGAGTTCCTCCAGCGAGCAGAACCGCGCCCAACAGCCCCACCAAGCCCGGAG	309	
Db	442	SGlyArgProGlyAsnPheLeuGlnInsArgProGluProThrAlaProProGluGlu	462	
QY	310	CTGCAGGTGCGCG-----GGCAACAACCCCGCAGCGAGCGCGCGCGCAGCGCCAGCGGCG	363	
Db	462	rPheArgPheGlyGluGlyLysThrThrProSerGlnLysGlnLuproIleAspLys---	481	
QY	364	ACCTGAATCTCCCCAGATCACCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGTGGGCG	423	
Db	482	GluLeu-----TyrProLeuAlaSerLeuArg-----	490	
QY	424	GGCCAGATCAAGGAGCCCTGCTGGACACCGGCGCCGACGACCCGTCTGGAGGAGATG	483	
Db	490	-----	490	
QY	484	AGCCTGCCCGCAAGTGGAAAGCCCAAGATGATCGCGGCGCATCGCGCGCTTTCATCAAGGTG	543	
Db	490	-----	490	
QY	544	CGCCAGTACGACAGATCCTGTAGTCAGATCTCGCGCAAGAGGCCCATCGGCACCGTGCTGTG	603	
Db	490	-----	490	

QY 1672 ACCTACAGATACACGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATG 1731  
 Db |||||  
 840 ThrTyrGlnIleTyrGlnGluProPheLysAsnLeuLysThrGlyLysTyrAlaArgMet 859  
 QY 1732 CGCACCGGCCACCAACAGCTGAAGCAGCTGACCGAGCGCGTGCAGAGATCCCATG 1791  
 Db |||||  
 860 ArgGlyAlaHisThrAsnAspValLysGlnLeuThrGluAlaValGlnLysIleThrThr 879  
 QY 1792 GAGACATCGTGATCGGGCAAGACCCCAAGTTCGCCCTGCCCATCCAGAGGAGACC 1851  
 Db |||||  
 880 GluSerIleValIleTyrGlyLysThrProLysPheLysLeuProIleGlnLysGluThr 899  
 QY 1852 TGGGAGACCTGTGTGACCGACTACTGCGAGCGACCTGCATCCCCGAGTGGGAGTTCGTG 1911  
 Db |||||  
 900 TrpGluThrTrpTrpThrGluTyrTrpGlnAlaThrTrpIleProGluTrpGluPheVal 919  
 QY 1912 AACACCCGCCCTCGTGAGCTGTGTACGAGCTGGAGAGGAGCCCATCATCGCGGCC 1971  
 Db |||||  
 920 AsnThrProLeuValLysLeuTrpTyrGlnLeuGluLysGluProIleValGlyAla 939  
 QY 1972 GAGACCTTCTAGCTGAGCGCGCCCAACCGCAGACCAAGATCGGCAAGSCCGCTAC 2031  
 Db |||||  
 940 GluThrPheTyrValAlaGlyAlaAlaAsnArgGluThrLysLeuGlyLysAlaGlyTyr 959  
 QY 2032 GTGACCGACCGGGCGCGCAGAGATCGTGAGCCTGACCGAGACCCACCAACCAAGAGACC 2091  
 Db |||||  
 960 ValThrAsnArgGlyArgGlnLysValValThrLeuThrAspThrThrAsnGlnLysThr 979  
 QY 2092 GAGCTGCAGGCCATCCAGCTGCGCCCTGCAGACACGCGCAGCGAGGTGAACATCGTGACC 2151  
 Db |||||  
 980 AlaLeuGlnAlaIleTyrLeuAlaLeuGlnAspSerGlyLeuGluValAsnIleValThr 999  
 QY 2152 GACAGCCAGTACGCCCTCGGCATCATCCAGCCCGCCAGCCGACAGCGAGCGAGCGAGCTG 2211  
 Db |||||  
 1000 AlaSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspGlnSerGluSerGluLeu 1019  
 QY 2212 GTGAACCAAGATCATCGACGAGCTGATCAAGAGGAGAGGTGTACTGAGCTGGGTGCC 2271  
 Db |||||  
 1020 ValAsnGlnIleIleGluGlnLeuIleLysLysGluLysValTyrLeuAlaTrpValPro 1039  
 QY 2272 GCCCAAGGGGATCGCGGCAACGAGCAGATCGACAAGCTGGTGAGCAAGGGCATCCGC 2331  
 Db |||||  
 1040 AlaHisLysGlyIleGlyAsnGluGlnValAspLysLeuValSerAlaGlyIleArg 1059  
 QY 2332 AAGGTGCTGCTTCCTGACCGCATCGAT 2358  
 Db |||||  
 1060 LysValLeuPheLeuAspGlyIleAsp 1068

Search completed: June 2, 2005, 05:29:26  
 Job time : 221.794 secs